Efficient Digital Encoding and Estimation of Noisy Signals

Haralabos Christos Papadopoulos

RLE Technical Report No. 623

May 1998

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Grant No. N00014-93-1-0686, the Air Force Office of Scientific Research under Grant No. F49620-96-1-0072, and in part by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement No. DAAL01-96-2-0002.

The Research Laboratory of Electronics
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139-4307

Efficient Digital Encoding and Estimation of Noisy Signals

by

Haralabos Christos Papadopoulos

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 1998, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In many applications in science and engineering one must rely on coarsely quantized and often unreliable noisy measurements in order to accurately and reliably estimate quantities of interest. This scenario arises, for instance, in distributed wireless sensor networks where measurements made at remote sensors need to be fused at a host site in order to decipher an information-bearing signal. Resources such as bandwidth, power, and hardware are usually limited and shared across the network. Consequently, each sensor may be severely constrained in the amount of information it can communicate to the host and the complexity of the processing it can perform.

In this thesis, we develop a versatile framework for designing low-complexity algorithms for efficient digital encoding of the measurements at each sensor, and for accurate signal estimation from these encodings at the host. We show that the use of a properly designed and often easily implemented control input added prior to signal quantization can significantly enhance overall system performance. In particular, efficient estimators can be constructed and used with optimized pseudo-noise, deterministic, and feedback-based control inputs, resulting in a hierarchy of practical systems with very attractive performance-complexity characteristics.

Thesis Supervisor: Gregory W. Wornell
Title: Cecil and Ida Green Associate Professor of Electrical Engineering

Acknowledgments

I wish to express my sincere gratitude to my thesis committee members, Professors Al Oppenheim and Mitch Trott, for their important contributions to this thesis. Al's involvement, in particular, has extended well beyond that of a thesis reader. I especially wish to thank him for our "uninhibited thinking" sessions, whose mark is clearly present many-fold throughout this document.

I also wish to acknowledge a number of DSPG and ATRP members, and other colleagues. First, I am greatly indebted to my "padrino" Vanni Aliberti, whose advice, support, and unique sense of humor have been invaluable throughout my doctoral program. I am also very thankful to Dr. Carl-Erik Sundberg, Steve Isabelle, John Apostolopoulos, and Andy Singer for being very supportive throughout this thesis. Thanks are also due to John Buck, the "ideal officemate" Warren Lam, and my "basketball student" Nick Laneman.

The generous financial support of the Office of Naval Research, the Air Force Office of Scientific Research, and the Army Research Laboratory is gratefully acknowledged.

I also wish to thank a number of friends that have withstood my mood and supported me many a time throughout this journey. In particular, I am deeply indebted to my close friend Thanos Siapas. His inspiring optimism and heartfelt support have been truly unmatched. Thanks are also due to the rest of the Tuna fans Alex Mantzaris, Nikos Giannakis, Peter Kofinas and, especially, Hayat Tazir, the "still in the closet" Tuna fans Ted Mihopoulos and Chris Hadjicostis, and of course, his Tunaship Bill Parcells, who has served as a prime example of a great mentor and has provided some very bright and memorable moments in the last few years.

I am deeply thankful to my two wonderful sisters Stella and Maria, my mother Eugene, and my "one of a kind" father Chris, for their boundless love, support, and encouragement.

Most of all, I am extremely grateful to my thesis supervisor Professor Greg Wornell. My appreciation for what Greg has done for me is truly hard to express in words. In the course of a lifetime, only a handful of people, if any, can have such a profound influence in shaping an individual, as Greg had with me. Greg will always serve as my role model as a teacher, researcher, and mentor. Truly, his supervision alone has made my doctoral program an experience that I will be very fond of for the rest of my life. Dedicating this work to Greg is just a small return to what he has given me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There is a wide range of applications in science and engineering where we wish to decipher signals from noisy measurements, and where system constraints force us to rely on a quantized or coarse description of those measurements. Representative examples include analog-to-digital (A/D) conversion, lossy compression, and decentralized data fusion. Indeed, in many data fusion problems the available resources place constraints in the type and the amount of data that can be exploited at the fusion center. Data fusion problems arise in a very broad and diverse range of applications, including distributed sensing for military applications [8], data-based management systems [2], target tracking and surveillance for robot navigation [22, 28] and radar applications [35], and medical imaging [9].

Recently, data fusion has attracted considerable attention in the context of distributed sensing problems, due to the continuing reduction in the cost of sensors and computation, and the performance improvements that inherently emanate from the use of multiple sensors [33]. Unlike classical multi-sensor fusion where the data collected by the sensors are communicated in full to a central processor, it is often desirable to perform some form of decentralized processing at the sensor before communicating the acquired information to the central processor in a condensed and often lossy form.

Various challenging signal detection and estimation problems have surfaced in such distributed sensing applications. Naturally, it is important to determine the extent to which decentralized preprocessing limits performance and to develop effective low-complexity methods for performing decentralized data fusion. As Hall *et al.* [19] show in the context of decentralized estimation, depending on the particular scenario, distributed data processing

may range from being optimal, in the sense that no loss in performance is incurred by simply communicating the local estimates computed at each sensor, to being catastrophic, in the sense that preprocessing at each sensor can completely destroy the underlying structure in the joint set of sensor measurements. Similar performance characteristics are exhibited in decentralized signal detection problems [5, 30]. Although for many important cases of practical interest decentralized signal detection and estimation methods have been formed for locally optimized processing at each sensor and subsequent efficient data fusion at the host (see [10, 34, 6, 30, 19, 11, 7, 13, 24] and the references therein), a number of real-time decentralized fusion problems are still largely unexplored.

In this thesis we focus on an important real-time decentralized fusion problem that arises in networks of distributed wireless sensors used for collecting macroscopic measurements. In particular, such networks are naturally suited for monitoring temporal variations in the average levels of environmental parameters. Representative examples include monitoring concentration levels in the atmosphere for detecting chemical or biological hazards, and measuring temperature fluctuations in the ocean surface for weather forecasting applications.

A block diagram of such a wireless sensor network is depicted in Fig. 1-1. In such a network, the local measurements made at each sensor must be communicated with minimal delay to a host over a wireless channel, where they must be effectively combined to decipher the information-bearing signal. Since bandwidth must often be shared across such a sensor network, the effective data rate at which each sensor can reliably communicate to the host over the wireless channel may be severely limited, often, to a few bits of information per each acquired sensor measurement. The need for power efficient design may also place constraints in the available processing complexity at each sensor, but usually not at the host, which typically possesses more processing power than each individual sensor. Depending upon bandwidth availability in these wireless networks, the host may or may not broadcast information back to the remote sensors, so as to improve the quality of the future sensor data it receives.

This type of problem may also arise in networks that are not wireless, but where the sensors are intrinsically limited by design. For instance, concentrations of chemical or biological agents are often computed by observing the color or the conformation of certain indicator/sensor molecules. In many of these cases, these sensor molecules exhibit only a

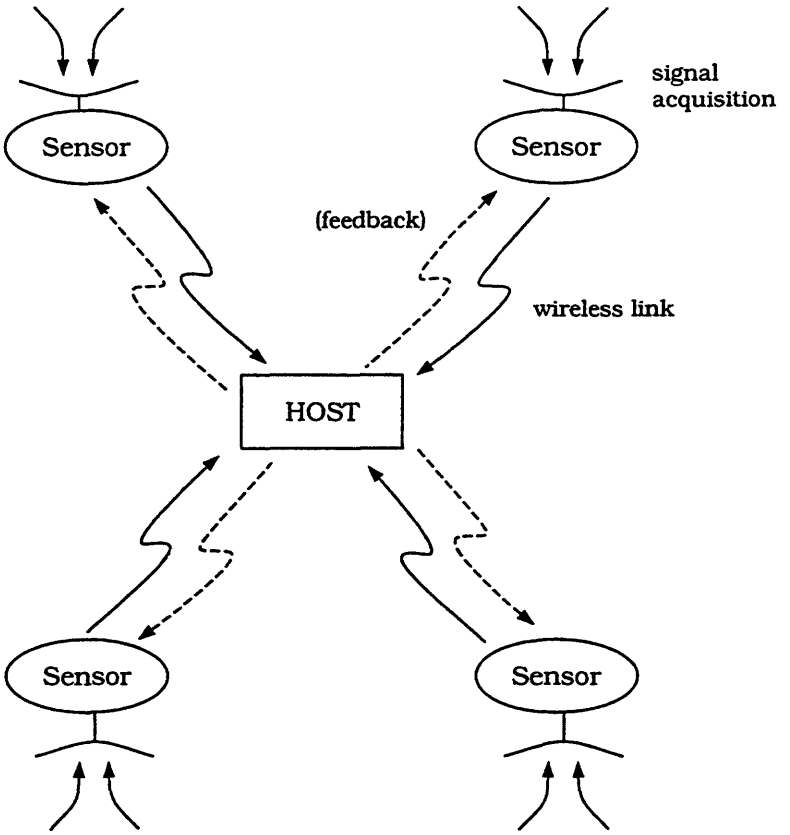
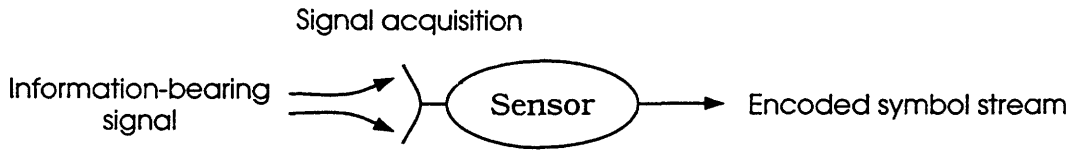


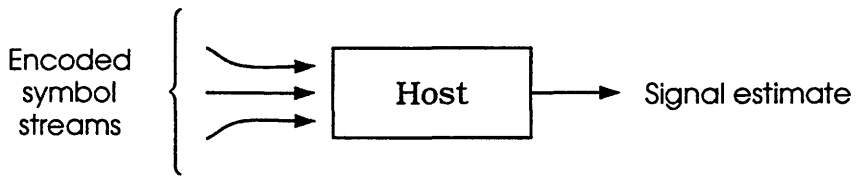
Figure 1-1: Block diagram of a wireless sensor network with bandwidth and power constraints.

finite set of possible outputs. In addition, there is often very limited flexibility in terms of affecting or biasing future outputs exhibited by these indicator molecules. Such networks of resolution-limited sensors are also employed by a number of biological systems for performing vital sensory tasks, suggesting that the type of processing performed by these systems somehow corresponds to an efficient use of resources [15, 16, 23]. For instance, it has been conjectured that certain types of crayfish enhance the ability of their crude sensory neurons to reliably detect weak signals sent by their predators by exploiting remarkably simple and, at first sight, counterintuitive pre-processing [16].

Various types of data fusion problems of the form depicted in Fig. 1-1 have been examined; in particular, the limitations in the amount of information that each sensor can communicate to the host are present in a number of decentralized detection problems [10, 32]. Another example that fits the same framework is what is referred to as the "CEO problem" [4], where a number of agents obtain noisy observations of a signal of interest and have to



(a) Encoding the noisy measurements of the information-bearing signal into a symbol stream at each sensor



(b) Estimation of information-bearing signal at the host from the encoded data streams

Figure 1-2: Framework for signal estimation from noisy measurements in sensor networks.

communicate this information to a CEO who can at most absorb R bits of information per second.

In this thesis, we focus on the problem of signal estimation in the context of sensor networks of the form depicted in Fig. 1-1, where system constraints limit the amount of information that each sensor can communicate to the host, and where there may also exist constraints in the amount of computation available at each sensor. It is very convenient to decompose this problem in the two stages shown in Fig. 1-2. First, as depicted in Fig. 1-2(a), at each sensor, the acquired noisy measurements of the information-bearing signal must be encoded into an efficient digital representation. Then, as shown in Fig. 1-2(b), the data streams from all sensors are to be effectively combined at the host in order to obtain an accurate signal estimate.

As we might expect, these two design stages of data encoding and signal estimation are very closely coupled. At each sensor, the measurements have to be efficiently encoded so as to enable the host to obtain an accurate signal estimate. Conversely, the host should exploit all the available information about the encoding strategy, and, in particular, when feedback is available, it may broadcast feedback information to the sensors so as to improve the quality of the future sensor encodings it receives. As we demonstrate in this thesis, the performance of the overall system strongly depends on the type of processing complexity constraints that are present at the sensor for encoding the sensor measurements.

1.1 Outline of the Thesis

In this thesis we develop a framework for designing computationally efficient algorithms for effective digital encoding of the measurements at each sensor, and for accurate signal estimation from these encodings at the host. In Chapters 3–5 we focus on the case where the information-bearing signal varies slowly enough that we may view it as static over the observation interval. We begin by examining in detail in Chapter 2 a class of low-complexity algorithms for encoding noisy measurements collected from a single sensor in the static case. Specifically, we consider encodings of the form of a suitably designed control input added prior to signal quantization. Depending on the amount of information that the estimator can exploit about the control input and the limitations in processing complexity at the encoder, a number of key encoding strategies and associated estimator structures are presented. For a number of scenarios of practical interest, we develop host efficient estimators that can be used with optimized control inputs at the sensor, resulting in a hierarchy of systems with very attractive performance-complexity characteristics.

In Chapter 3, we develop a number of important extensions of the systems developed in Chapter 2 which can be useful in the context of networks of sensors. We first develop optimized multi-sensor extensions of the single-sensor encoding and estimation strategies for all the scenarios considered in Chapter 2. These systems have a number of important potential applications, especially in the context of distributed sensing networks where there are physical limitations in the sensor design, or bandwidth and power constraints. We also develop extensions of these encoders and estimators for scenarios where prior information about the information-bearing signal is available. In addition, we consider the case where the sensor noise power level is also unknown, and develop the performance limits and the associated extensions of the encoders and estimators for all the scenarios considered in Chapter 2.

In Chapter 4, we consider more general encoding strategies for the static signal case. These encoders require more complex processing than the encoders employing quantizer bias control of Chapters 2–3 and are therefore attractive when there are less stringent complexity constraints at the encoder. As we demonstrate, we can develop refinable encoding and estimation strategies which asymptotically achieve the best possible performance based on the original sensor measurements. In this sense, we show that using a suitably designed

digitized description of the acquired noisy measurements does not incur any performance loss in signal estimation.

In Chapter 5, we consider a number of extensions of the static-case encoding strategies which encompass a broad class of time-varying signals. In particular, in the case that the same time-varying signal is observed at each sensor, a rich class of algorithms can be designed for measurement encoding by exploiting the encoding principles for the static problem. In fact, these methods can be applied in the context of a large class of signal models, namely, signals that can be characterized by conventional state-space models. As we also show, for such information-bearing signals we can also design effective estimation algorithms which are based on extensions of conventional Kalman filtering solutions.

Finally, a summary of the main contributions of this thesis is given in Chapter 6, along with a representative collection of potentially interesting directions for future research that are suggested by this work.

Chapter 2

Encoding from Noisy Measurements via Quantizer Bias Control: Static Case

In developing methods for overcoming the power/bandwidth constraints that may arise across a sensor network, or the dynamic range and resolution constraints at each sensor, it is instructive to first examine the single-sensor problem. In fact, this special case captures many of the key design and performance issues that arise in the context of networks of sensors. The block diagram corresponding to a single sensor is shown in Fig. 2-1, where $A[n]$ denotes the information-bearing signal, $v[n]$ represents sensor noise, $s[n]$ denotes the sensor measurement sequence, and $y[n]$ denotes the sequence of M-ary symbols encoded at the sensor and used at the host to obtain a signal estimate $\hat{A}[n]$. Consistent with the system constraints, throughout the thesis, we focus on developing algorithms that generate encoded sequences whose *average* encoding rate does not exceed one M-ary symbol per available sensor measurement. The task is then to design the encoder at the sensor and the associated estimator from the encodings at the host so as to optimize the host estimate quality.

To illustrate some of the key issues that may arise in the encoder design, it is insightful to consider the static case, *i.e.*, the case where the signal $A[n]$ is varying slowly enough that we may view it as static over the observation interval. Given a fixed time instant N, we can easily devise a method for efficiently encoding the N sensor measurements

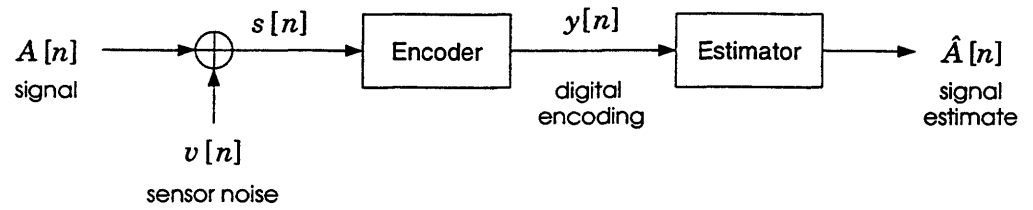


Figure 2-1: Block diagram of encoding the noisy measurements at the sensor and signal estimation from these encodings at the host.

$s[1], s[2], \cdots, s[N]$, into a sequence of N M-ary symbols $y[1], y[2], \cdots, y[N]$ provided N is large. Specifically, consider the following algorithm:

At the sensor:

(i) compute an estimate of the static information-bearing signal using the N sensor measurements;

(ii) quantize the estimate using a uniform quantizer with M^N quantization levels;

(iii) communicate to the host the quantized level by means of the N M-ary symbols $y[1], y[2], \cdots, y[N]$.

At the host:

(i) reconstruct the "quantized" estimate using $y[1], y[2], \cdots, y[N]$.

Clearly, since the number of available quantization levels in step (ii) of the encoder grows exponentially with the number of available observations N, the error between the "quantized" estimate used at the host and the original sensor estimate produced in step (i) of the encoder (*i.e.*, the estimate prior to quantization) decays exponentially fast with N.

A major disadvantage of such an encoding scheme, however, is that it is not refinable, namely it provides an one-shot description; no encodings are available to the host for forming estimates before time N, and no encodings are available after time N to further refine the quality of the host estimate. Furthermore, this encoding scheme assumes that there is absolute freedom in designing the M^N-level quantizer. However, this is often not the case such as in problems where the sensors are intrinsically limited by design. For these reasons, in this thesis we instead focus on designing refinable encoding strategies.

One of simplest refinable encoding strategies that can be constructed consists of quantizing each noisy measurement at the sensor by means of an M-level quantizer. As we

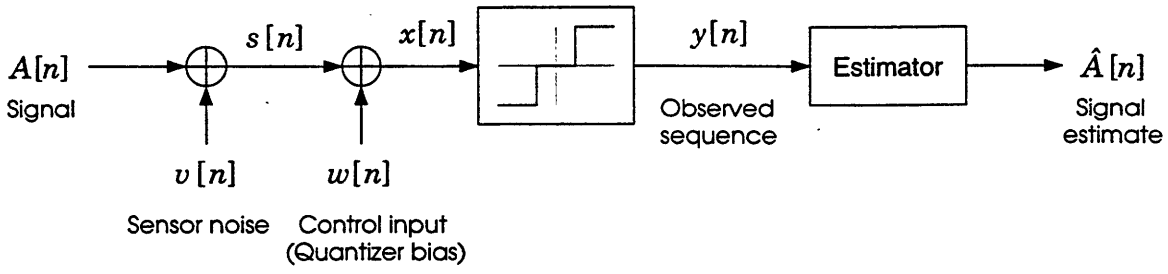


Figure 2-2: Signal estimation based on digital encodings which are generated by adding a suitably designed control input prior to signal quantization.

show in this chapter, however, this simple encoding scheme can have very poor performance characteristics, in terms of overcoming the power/bandwidth constraints across the network, or the dynamic range and resolution constraints at the sensor. As a means for improving the effective digital encoding we may consider the use of a control input added to the information-bearing signal prior to quantization at the sensor. The block diagram corresponding to a single sensor in the context of such an encoding scheme is shown in Fig. 2-2, where $w[n]$ is a control input, and, as in Fig. 2-1, $A[n]$ denotes the information-bearing signal, $v[n]$ represents sensor noise, and $y[n]$ denotes the quantized signal that is sent to the central site.

In this chapter we focus on the static case of the estimation problem depicted in Fig. 2-2 in which $A[n] = A$, *i.e.*, we examine the problem of estimating a noise-corrupted unknown parameter A via quantized observations. This case reveals several key features of signal estimation from quantized observations obtained via a network of sensor encoders, each comprising a control input and a quantizer; in Chapter 5 we develop extensions of our analysis corresponding to the dynamic scenario where $A[n]$ is time-varying.

Several basic variations of the encoding and estimation problem depicted in Fig. 2-2 can arise in practice, which differ in the amount of information about the control input that is available for estimation and the associated freedom (or available encoding complexity) in the control input selection. In this chapter we develop effective control input selection strategies and associated estimators for all these different scenarios. In particular, for pseudo-noise control inputs whose statistical characterization alone is exploited at the receiver, we show that there is an optimal power level for minimizing the mean-square estimation error (MSE). The existence of a non-zero optimal pseudo-noise power level reveals strong connections to the phenomenon of stochastic resonance, which is encountered in a number of physical

nonlinear systems where thresholding occurs and where noise is often exploited for signal enhancement [3, 16, 18]. Performance can be further enhanced if detailed knowledge of the applied control waveform is exploited at the receiver. In this scenario, we develop methods for judiciously selecting the control input from a suitable class of periodic waveforms for any given system. Finally, for scenarios where feedback from the quantized output to the control input is available, we show that, when combined with suitably designed receivers, these signal quantizers come within a small loss of the quantizer-free performance.[1] In the process we develop a framework for constructing the control input from past observations and design computationally efficient estimators that effectively optimize performance in terms of MSE.

The outline of this chapter is as follows. In Section 2.1 we describe the static-case estimation problem associated with the system depicted in Fig. 2-2. In Section 2.2 we develop the estimation performance limits for a number of important scenarios. In Section 2.3 we design control inputs and associated estimators for each of these distinct scenarios, which achieve the performance limits developed in Section 2.2. Finally, in Section 3.1 we examine a network generalization of the scenario depicted in Fig. 2-2, in which signal estimation is based on quantized observations collected from multiple sensors.

2.1 System Model

As outlined above, in this chapter we consider the problem of estimating an unknown parameter A from observation of

$$y[n] = F(A + v[n] + w[n]) \qquad n = 1, 2, \cdots, N \,, \tag{2.1}$$

where the sensor noise $v[n]$ is an independent identically distributed (IID) process, $w[n]$ is a control input, and the function $F(\cdot)$ is an M-level quantizer, with the quantized output

[1] Although the feedback loop can be entirely implemented at the sensor, sensor complexity is reduced by having the feedback information come from the central site. This is especially appealing in wireless networks where power resources at the central site are often such that there is plenty of effective bandwidth available for broadcasting high-resolution control information.

$y[n]$ taking M distinct values $Y_1, Y_2, \cdots, Y_M$, *i.e.*,

$$F(x) = \begin{cases} Y_i & \text{if } X_{i-1} \leq x < X_i, \text{ for } 2 \leq i \leq M \\ Y_1 & \text{otherwise} \end{cases} , \tag{2.2a}$$

where $X_0 = -\infty$ and $X_M = \infty$. Without loss of generality, we assume that the quantizer levels are uniformly spaced, *i.e.*,

$$Y_i = -(M+1) + 2\,i\,, \qquad i = 1, 2, \cdots, M\ ; \tag{2.2b}$$

any other set of distinct quantization levels is equivalent to (2.2b) in the sense that the two sets are related by means of an invertible transformation. We also define the intermediate sequence

$$s[n] \stackrel{\Delta}{=} A + v[n] = A + \sigma_v\, \tilde{v}[n]\ . \tag{2.3}$$

We will frequently be interested in a measure of predicted performance for a family of sensor noises parameterized by σ_v in (2.3), arising from scaling an IID noise sequence $\tilde{v}[n]$. We use the notation $p_z(\cdot)$ to denote the probability density function (PDF) of any sample of an IID sequence $z[n]$, and $C_z(\cdot)$ to denote one minus the corresponding cumulative distribution, *i.e.*,

$$C_z(x) = \int_x^\infty p_z(t)\ dt\ .$$

We shall refer to an IID noise process as *admissible* if the associated PDF is non-zero and smooth (*i.e.*, C^1) almost everywhere. Throughout this chapter, we assume that all noise processes are admissible, including $v[n]$ as well as $w[n]$, when $w[n]$ is viewed as a pseudo-noise process. Furthermore, when referring to a Gaussian process we assume it is IID and zero-mean, unless we specify otherwise.

2.2 Performance Limits for Controllers with Quantizer Bias Control

In this section we quantify the performance degradation that results from estimating A based on observation of $y[n]$ instead of $s[n]$. We first introduce the concept of *information loss*, which we use as a figure of merit to design quantizer systems and evaluate the associated estimators. We then present a brief preview of performance limits based on this notion for a number of important scenarios and finally develop these performance limits in Sections 2.2.1–2.2.3.

We define the information loss for a quantizer system as the ratio of the Cramér-Rao bounds for unbiased estimates of the parameter A obtained via $y[n]$ and $s[n]$, respectively, *i.e.*,

$$\mathcal{L}(A) \triangleq \frac{\mathcal{B}\left(A;\ \mathbf{y}^N\right)}{\mathcal{B}\left(A;\ \mathbf{s}^N\right)}\ , \tag{2.4}$$

where $\mathcal{B}\left(A;\ \mathbf{y}^N\right)$ is the Cramér-Rao bound [31] for unbiased estimation of A from[2]

$$\mathbf{y}^N \triangleq \begin{bmatrix} y[1] & y[2] & \cdots & y[N] \end{bmatrix}^T\ , \tag{2.5}$$

where $y[n]$ is given by (2.1), and where $\mathcal{B}\left(A;\ \mathbf{s}^N\right)$ and $\mathbf{s}^N$ are defined similarly. We often consider the information loss (2.4) in dB, (*i.e.*, $10\log_{10}\mathcal{L}(A)$); it represents the additional MSE in dB that arises from observing $y[n]$ instead of $s[n]$ in the context of efficient estimation of A. From this perspective, better systems achieve smaller information loss over the range of parameter values of interest.

Taking into account the inherent dynamic range limitations of these signal quantizers, we assume that the unknown parameter A takes values in the range $(-\Delta,\ \Delta)$, with Δ assumed to be known. Often, the degradation of the estimation quality is conveniently characterized in terms of the ratio $\chi = \Delta/\sigma_v$, which we may view as a measure of peak-signal-to-noise ratio (peak SNR).

Worst-case performance is used to characterize the overall system. Accordingly, we

[2]The use of the term information loss follows from the fact that (2.4) also equals the inverse of the ratio of the associated Fisher information quantities.

define the worst-case Cramér-Rao bound and worst-case information loss via

$$\mathcal{B}_{\max}(\Delta) \triangleq \sup_{|A|<\Delta} \mathcal{B}\left(A;\ \mathbf{y}^N\right), \tag{2.6}$$

and

$$\mathcal{L}_{\max}(\Delta) \triangleq \sup_{|A|<\Delta} \mathcal{L}(A), \tag{2.7}$$

respectively. Both the worst-case Cramér-Rao bound and the worst-case information loss are functions of other system parameters, such as σ_v and $F(\cdot)$, the dependence on which is suppressed for convenience in the above definitions.

As a consequence of the linear model (2.3), the Cramér-Rao bound $\mathcal{B}\left(A;\ \mathbf{s}^N\right)$ is independent of the parameter value A, *i.e.*, $\mathcal{B}\left(A;\ \mathbf{s}^N\right) = \mathcal{B}\left(0;\ \mathbf{s}^N\right)$ for any A. Furthermore, the bound $\mathcal{B}\left(A;\ \mathbf{s}^N\right)$ is proportional to σ_v^2; by letting $\tilde{s}[n] = A + \tilde{v}[n]$ and using (2.3), we obtain

$$\mathcal{B}\left(A;\ \mathbf{s}^N\right) = \sigma_v^2\, \mathcal{B}\left(0;\ \tilde{s}\right)/N, \tag{2.8}$$

where $\mathcal{B}\left(0;\ \tilde{s}\right)$ denotes the Cramér-Rao bound for estimating A based on any one sample of the IID sequence $\tilde{s}[n]$. Hence, since $\mathcal{B}\left(A;\ \mathbf{s}^N\right)$ from (2.8) is independent of A, both $\mathcal{B}_{\max}(\Delta)$ and $\mathcal{L}_{\max}(\Delta)$ can be used interchangeably as figures of merit for assessing the performance of quantizer systems.

Table 2.1 summarizes the performance limits for a number of important scenarios. As we show in this chapter, in any of these scenarios the worst-case information loss can be conveniently characterized as a function of peak SNR χ. According to Table 2.1, pseudo-noise control inputs with properly chosen power levels provide performance improvements over control-free systems in any admissible noise. Specifically, for pseudo-noise control inputs the control input can be designed so that the worst-case information loss grows only quadratically with χ, while it always grows faster than quadratically in the control-free case. For scenarios where the control input is known for estimation, the associated worst-case loss can be made to grow as slow as χ with proper control input selection. Finally, if feedback from the quantized output to the control input is available and properly used, a fixed small information loss, which does not grow with increasing χ, can be achieved. In the remainder

Control Input	Order of growth of information loss	
	Gaussian case	General case
Control-free case	$e^{\chi^2/2}$	$> \chi^2$
Pseudo-noise (known statistics)	χ^2	χ^2
Known input	χ	χ
Feedback-controlled input	1	1

Table 2.1: Order of growth of worst-case information loss as a function of peak SNR $\chi = \Delta/\sigma_v$ for large χ and for any M-level quantizer. The quantity Δ denotes the dynamic range of the unknown parameter, and σ_v is the sensor noise power level. The Gaussian case refers to Gaussian sensor noise of variance σ_v^2. The general case refers to any admissible sensor noise.

of Section 2.2 we develop the performance limits shown in Table 2.1, while in Section 2.3 we develop control selection methods and associated estimators that achieve these limits.

2.2.1 Pseudo-noise Control Inputs

In this section we consider signal quantizers with control inputs $w[n]$ that correspond to sample paths of an IID process, independent of the sensor noise process $v[n]$, and determine the performance limits in estimating the unknown parameter A based on observation of $\mathbf{y}^N$ from (2.5), by simply exploiting the statistical characterization of $w[n]$ at the receiver. In general, we may consider pseudo-noise control inputs that are parameterized by means of the scale parameter σ_w, *i.e.*, $w[n] = \sigma_w \tilde{w}[n]$, where $\tilde{w}[n]$ is an admissible IID noise sequence with PDF $p_{\tilde{w}}(\cdot)$. Our goal is to select the pseudo-noise level σ_w so as to optimize performance in terms of the associated worst-case information loss.[3]

The Cramér-Rao bound for all unbiased estimates of the parameter A based on observation of the vector $\mathbf{y}^N$ is defined as [31]

$$\mathcal{B}(A;\ \mathbf{y}^N) = -\left(E\left[\frac{\partial^2 \ln P(\mathbf{y}^N; A)}{\partial A^2}\right]\right)^{-1},$$

where $P(\mathbf{y}^N; A)$ is the associated likelihood function, denoting the probability that the particular vector $\mathbf{y}^N$ is observed from (2.1) given that the unknown parameter takes the

[3]The scaling factor σ_w is a measure of the strength of the noise process $w[n]$. For cases where the noise variance exists, σ_w^2 denotes the power of the pseudo-noise signal to within a scaling.

value A. In particular, the log-likelihood function satisfies

$$\ln P(\mathbf{y}^N; A) = \sum_{i=1}^{M} \mathcal{K}_{Y_i}\left(\mathbf{y}^N\right) \ln \Pr(y[n] = Y_i; A) \tag{2.9}$$

where $\mathcal{K}_{Y_i}\left(\mathbf{y}^N\right)$ denotes the number of entries in $\mathbf{y}^N$ that are equal to Y_i. Since

$$\alpha[n] = v[n] + w[n] \tag{2.10}$$

is an IID sequence, $\mathcal{B}\left(A;\ \mathbf{y}^N\right)$ satisfies the condition

$$\mathcal{B}\left(A;\ \mathbf{y}^N\right) = \frac{1}{N}\mathcal{B}\left(A;\ y\right) , \tag{2.11}$$

where $\mathcal{B}\left(A;\ y\right)$ corresponds to the Cramér-Rao bound for estimating A based on any one sample of the IID sequence $y[n]$. Finally, by taking the second partial derivative of (2.9) with respect to A followed by an expectation, we obtain

$$\mathcal{B}\left(A;\ y\right) = \left(\sum_{i=1}^{M} \frac{\left[p_\alpha\left(X_{i-1} - A\right) - p_\alpha\left(X_i - A\right)\right]^2}{C_\alpha\left(X_{i-1} - A\right) - C_\alpha\left(X_i - A\right)}\right)^{-1} . \tag{2.12}$$

For the system corresponding to the symmetric two-level quantizer ($M = 2$, $X_1 = 0$), *i.e.*,

$$F(x) = \operatorname{sgn}\ x , \tag{2.13}$$

the Cramér-Rao bound (2.12) reduces to

$$\mathcal{B}\left(A;\ y\right) = C_\alpha\left(-A\right)\left[1 - C_\alpha\left(-A\right)\right]\left[p_\alpha\left(-A\right)\right]^{-2} . \tag{2.14}$$

When, in addition, the PDF $p_\alpha(\cdot)$ is an even function of its argument, (2.14) further specializes to

$$\mathcal{B}\left(A;\ y\right) = \mathcal{B}\left(-A;\ y\right) = C_\alpha\left(-A\right) C_\alpha\left(A\right)\left[p_\alpha\left(A\right)\right]^{-2} . \tag{2.15}$$

We consider the special case where $v[n]$ and $w[n]$ are IID Gaussian processes and $F(\cdot)$ is the symmetric two-level quantizer, and determine the pseudo-noise level that minimizes

the worst-case information loss. We then consider the general case, *i.e.*, the case $M \geq 2$ where $v[n]$ and $w[n]$ are any IID noise processes.

Special Case: Gaussian Noises and $M = 2$

For the system $M = 2$ where $v[n]$ and $w[n]$ are independent IID Gaussian noise sequences with variances σ_v^2 and σ_w^2 respectively, the Cramér-Rao bound (2.15) reduces to

$$\mathcal{B}(A;\ y) = 2\,\pi\,\sigma_\alpha^2\,Q\left(\frac{A}{\sigma_\alpha}\right)\,Q\left(-\frac{A}{\sigma_\alpha}\right)\,\exp\left(\frac{A^2}{\sigma_\alpha^2}\right)\,, \tag{2.16}$$

where $\sigma_\alpha = \sqrt{\sigma_w^2 + \sigma_v^2}$, and $Q(x) = \int_x^\infty (1/\sqrt{2\pi})\,e^{-t^2/2}\,dt$. Fig. 2-3 depicts the associated information loss (2.4) as a function of A for $\Delta = 1$, $\sigma_v = 0.1$ and various σ_w levels. Observation of Fig. 2-3 reveals several key characteristics of this type of quantizer-based processing. Specifically, in this Gaussian sensor noise scenario the minimum achievable information loss occurs for $A = 0$ and $\sigma_w = 0$ and equals $10\log 10(\pi/2) \approx 2$ dB. In addition, for any pseudo-noise power level σ_w the information loss is an increasing function of $|A|$. This property is shared by many other common noises, such as the Laplacian and the Cauchy. More important, as the figure reveals, proper use of pseudo-noise ($\sigma_w \neq 0$) can have a major impact on performance in terms of reducing the associated worst-case information loss.

The sensitivity of performance with respect to the optimal pseudo-noise power level is examined in Fig. 2-4 for the Gaussian noise scenario. In particular, the figure depicts the additional worst-case information loss (in dB) that arises from suboptimal selection of the pseudo-noise power level. Since the encoding performance for the optimally selected pseudo-noise power level is used as a reference, the additional worst-case information loss for the optimal pseudo-noise encoder equals zero dB. From the figure we see that the optimal aggregate noise level is well approximated by

$$\sigma_\alpha^{\text{opt}} \approx \frac{2}{\pi}\,\Delta\,, \tag{2.17}$$

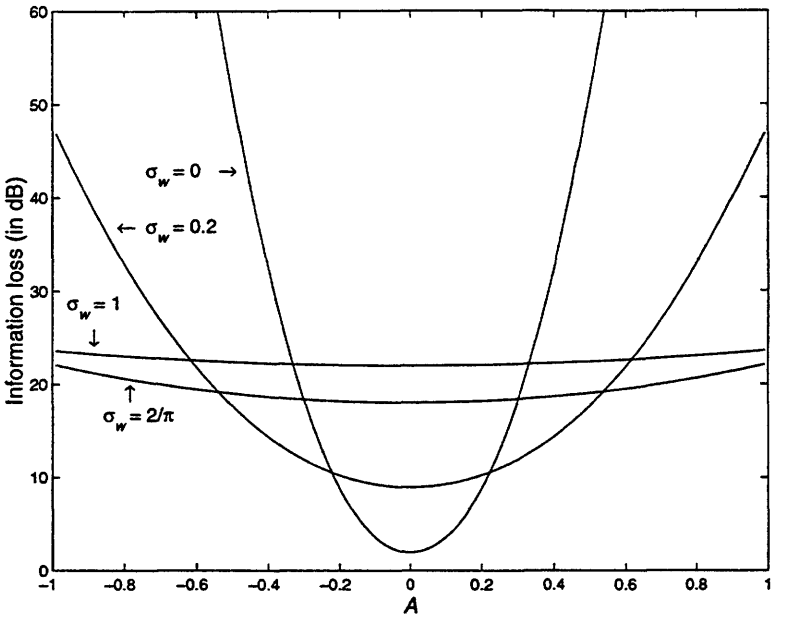


Figure 2-3: Information loss for a system comprising a two-level quantizer and an IID Gaussian pseudo-noise control input, for various pseudo-noise power levels σ_w. The sensor noise is IID Gaussian with variance $\sigma_v^2 = 0.01$.

so that the optimal pseudo-noise level satisfies

$$\sigma_w^{\text{opt}} = \begin{cases} \sqrt{\left(\sigma_\alpha^{\text{opt}}\right)^2 - \sigma_v^2} & \text{if } \sigma_v < \sigma_\alpha^{\text{opt}} \\ 0 & \text{otherwise} \end{cases} . \tag{2.18}$$

If $\sigma_v \ll \Delta$ (high SNR), Fig. 2-4 reveals that for the fairly wide range of pseudo-noise levels

$$\frac{5}{8}\,\sigma_\alpha^{\text{opt}} \leq \sigma_w \leq 2\,\sigma_\alpha^{\text{opt}} ,$$

the associated performance is inferior to that corresponding to the optimal pseudo-noise level by less than 3 dB. However, the performance degrades rapidly as the pseudo-noise level is reduced beyond $5\,\sigma_\alpha^{\text{opt}}/8$. For instance, for $\sigma_w = \sigma_\alpha^{\text{opt}}/3$, there is nearly 30 dB of additional loss incurred by the suboptimal selection of the pseudo-noise level.

The information loss associated with the optimal pseudo-noise level corresponds to the best achievable performance by a particular family of pseudo-noise sources—in this particular example the family of zero-mean normal distributions. For the optimal choice of σ_w in (2.18), the worst-case information loss can be completely characterized by means of peak

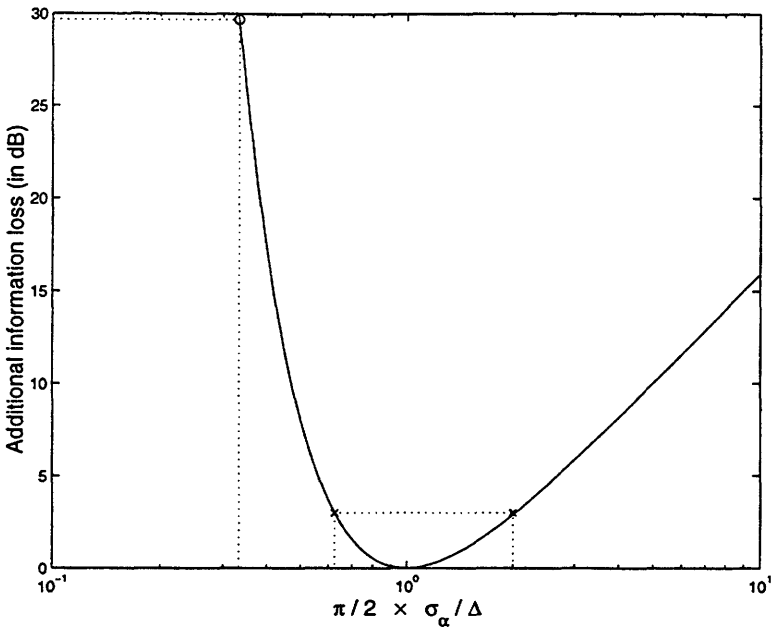


Figure 2-4: Additional worst-case information loss (solid) due to suboptimal pseudo-noise level selection for a two-level quantizer. The net noise sequence $\alpha[n] = v[n] + w[n]$ is Gaussian with variance σ_α^2. The "×" marks depict the additional information loss for net noise levels $5/8\,\sigma_\alpha^{\text{opt}}$ and $2\,\sigma_\alpha^{\text{opt}}$. The "o" mark depicts the additional information loss at $\sigma_\alpha^{\text{opt}}/3$.

SNR χ. In particular, by using (2.17)–(2.18) with (2.16) in (2.4) we obtain the optimal worst-case information loss for the Gaussian scenario with pseudo-noise control, namely,

$$\mathcal{L}_{\max}^{\text{pn}}(\chi) = \begin{cases} 2\pi Q(\chi) Q(-\chi)\, e^{\chi^2} & \text{if } 0 \le \chi \le \frac{\pi}{2} \\ \frac{8}{\pi} Q\left(\frac{\pi}{2}\right) Q\left(-\frac{\pi}{2}\right) e^{\frac{\pi^2}{4}} \chi^2 & \text{if } \frac{\pi}{2} \le \chi \end{cases}, \tag{2.19}$$

where we indicate explicitly that in this case the worst-case information loss is a function of χ.

As (2.19) reveals, for estimation in Gaussian noise via a two-level quantizer system, the worst-case information loss can be made to grow quadratically with peak SNR by judicious selection of a Gaussian pseudo-noise control input. For comparison, the worst-case information loss in the absence of control input grows exponentially with peak SNR. In particular, by substituting $\mathcal{B}(A;\, y)$ from (2.16) in (2.7), we obtain

$$\mathcal{L}_{\max}^{\text{free}}(\chi) = 2\pi Q(\chi)\, Q(-\chi)\, e^{\chi^2}, \tag{2.20}$$

which is proportional to $\exp(\chi^2/2)$ for large χ. .The results in (2.19)–(2.20) extend to quantizers with $M > 2$, *i.e.*, the worst-case information loss grows as $\exp(\chi^2/2)$ for control-free systems, while it can be made to grow as χ^2 for appropriately chosen Gaussian pseudo-noise control inputs.

General Case: Arbitrary Noises and $M \geq 2$

As we show next, proper use of a pseudo-noise control input $w[n]$ can improve performance over the control-free system in any (admissible) sensor noise $v[n]$ and for any M-level quantizer. Substituting (2.8) and (2.11) in (2.4) reveals that the associated information loss is independent of N. Thus, we may focus on the case $N = 1$ without any loss of generality. We next use $\mathcal{B}_{\max}(\Delta; \sigma_v, \sigma_w)$ to denote the worst-case Cramér-Rao bound (2.6), in order to make its dependence on σ_v and σ_w explicit. Since $\tilde{v}[n]$ is an admissible process, the Cramér-Rao bound (2.12) is continuous in the σ_v variable, and so is $\mathcal{B}_{\max}(\Delta; \sigma_v, \sigma_w)$. Thus, given any fixed $\sigma_w > 0$ and Δ, for small enough σ_v we have

$$\mathcal{B}_{\max}(\Delta; \sigma_v, \sigma_w) \approx \mathcal{B}_{\max}(\Delta; 0, \sigma_w) \ . \tag{2.21}$$

Substitution of (2.21) and (2.8) in (2.7) reveals that $\mathcal{L}^{\text{pn}}_{\max}(\chi) \sim \chi^2$ is achievable for large χ . Furthermore, since $\mathcal{B}_{\max}(\Delta; \sigma_v, \sigma_w)$ is also continuous in σ_w, for any $F(\cdot)$ with fixed $M < \infty$

$$\inf_{\sigma_w \in [0, \infty)} \mathcal{B}_{\max}(\Delta; 0, \sigma_w) > 0 \ , \tag{2.22}$$

which in conjunction with (2.8) and (2.21) implies that the worst-case information loss can not be made to grow slower than χ^2 for pseudo-noise control inputs. Therefore, at high peak SNR the worst-case information loss for pseudo-noise control inputs $\mathcal{L}^{\text{pn}}_{\max}(\chi)$ grows quadratically with peak SNR for pseudo-noise control inputs. In general, the sensor noise level may be fixed, in which case we are interested in selecting the pseudo-noise level σ_w as a function of the dynamic range Δ so as to minimize the worst-case information loss. From (2.21)–(2.22) the optimal worst-case information loss rate can be achieved by selecting $\sigma_w = \lambda \Delta$ for some $\lambda > 0$. This is in agreement with our conclusions for the Gaussian scenario in the special case $M = 2$, as (2.17)–(2.19) clearly demonstrate. For comparison, in App. A.1 we show that for control-free systems corresponding to $F(\cdot)$ in (2.2) and for

any sensor noise the worst-case information loss $\mathcal{L}_{\max}^{\text{free}}(\chi)$ grows faster than χ^2 for large χ. Remarkably, pseudo-noise control inputs with appropriately selected power levels provide performance improvements over the control-free systems for any sensor noise at high peak SNR.

2.2.2 Known Control Inputs

We next develop performance limits for scenarios where the estimator can exploit detailed knowledge of a suitably designed control waveform. In particular, we determine the minimum possible growth rate of the worst-case information loss as a function of χ, and develop control input selection strategies that achieve the minimum possible rate.

The Cramér-Rao bound for unbiased estimates of A based on $\mathbf{y}^N$ and given knowledge of the associated N samples of $w[n]$ is denoted by $\mathcal{B}\left(A;\ \mathbf{y}^N,\ \mathbf{w}^N\right)$ and satisfies

$$\begin{aligned}\mathcal{B}\left(A;\ \mathbf{y}^N,\ \mathbf{w}^N\right) &= -\left(E\left[\frac{\partial^2 \ln P(\mathbf{y}^N; A, \mathbf{w}^N)}{\partial A^2}\right]\right)^{-1} \\ &= \left[\sum_{n=1}^{N}\left[\mathcal{B}\left(A + w[n];\ y\right)\right]^{-1}\right]^{-1}, \end{aligned} \tag{2.23}$$

where $\mathcal{B}\left(A;\ y\right)$ is given by (2.12), with α replaced by v, and where $P(\mathbf{y}^N; A, \mathbf{w}^N)$ denotes the associated likelihood function. As expected, the associated worst-case Cramér-Rao bound and worst-case information loss are functions of the control waveform $\mathbf{w}^N$. In App. A.2 we show that, for any known control waveform selection strategy, the worst-case information loss associated with any M-level signal quantizer grows at least as fast as χ for any sensor noise distribution. This includes the optimal scheme, which selects the waveform $w[n]$ that results in minimizing the worst-case information loss for any given set $\{\Delta,\ \sigma_v,\ p_{\bar{v}}\left(\cdot\right),\ F(\cdot)\}$.

Classes of periodic waveforms parameterized by the period K are appealing candidates for known control inputs, since they are easy to construct and can be chosen so that the worst-case information loss grows at the minimum possible rate. In constructing these classes of periodic waveforms, we use as a figure of merit the worst-case information loss for $N \to \infty$; extensions to the finite N case are developed in App. A.2. From (2.23), the Cramér-Rao bound for estimating A based on $\mathbf{y}^N$, where N is a multiple of the period K,

is given by,

$$\mathcal{B}\left(A;\ \mathbf{y}^N,\ \mathbf{w}^N\right) \ = \ \frac{1}{N}\,\frac{K}{\sum_{n=1}^{K}\left[\mathcal{B}\left(A+w[n];\ y\right)\right]^{-1}}\ . \tag{2.24}$$

As we show next, in order to achieve the minimum possible growth rate it suffices to select $w[n]$ from properly constructed K-periodic classes for which there is an one-to-one correspondence between each element in the class and the period K. Optimal selection of the control input in this case is equivalent to selecting the period K that minimizes the associated worst-case information loss, or equivalently, the worst-case Cramér-Rao bound from (2.24)

$$K_{\text{opt}}\left(\Delta,\sigma_v\right) \ \stackrel{\Delta}{=}\ \arg\min_{K}\ \sup_{A\ \in\ (-\Delta,\,\Delta)}\ \frac{K}{\sum_{n=1}^{K}\left[\mathcal{B}\left(A+w[n];\ y\right)\right]^{-1}} \tag{2.25}$$

where $\mathcal{B}\left(A;\ y\right)$ is given by (2.12) with α replaced by v. We next develop a framework for selecting the control waveform from properly constructed classes of K-periodic waveforms for the case $M = 2$, which results in achieving the optimal growth rate of worst-case information loss. Then, we extend our framework to quantizers with $M > 2$.

Optimized Periodic Waveforms for Signal Quantizers with $M = 2$

The construction of the elements of the K-periodic class in the case $M = 2$ is based on the observation that in the control-free scenario the worst-case information loss grows with Δ for fixed σ_v. This observation suggests that the information loss is typically largest for parameter values that are furthest from the quantizer threshold. This is strictly true, for instance, for Gaussian sensor noise, since $\mathcal{B}\left(A;\ y\right)$ in (2.16) is an increasing function of $|A|$. Since our objective is to optimize over the worst-case performance, a potentially appealing strategy is to construct the K-periodic waveform $w[n]$ so as to minimize the largest distance between any A in $(-\Delta,\,\Delta)$ and the closest *effective* quantizer threshold. For this reason, we consider K-periodic control inputs, which have the form of the sawtooth waveform

$$w[n] = \delta_w\left(-\frac{K-1}{2} + n \bmod K\right)\ , \tag{2.26}$$

where the effective spacing between thresholds is given by $\delta_w = 2\,\Delta/(K-1)$. The net effect of the periodic control input (2.26) and the symmetric two-level quantizer (2.13) is

equivalent to a two-level quantizer with a periodically time-varying threshold; it is important to observe that the time-varying quantizer threshold comes within at least $\delta_w/2$ of any possible parameter value once every K samples.

For the system with $F(\cdot)$ given by (2.13) and $w[n]$ given by (2.26), the optimal period K_{opt} is completely characterized by means of peak SNR χ; using (2.14) in (2.25) reveals that K_{opt} satisfies $K_{\text{opt}}(\Delta, \sigma_v) = K_{\text{opt}}(\mu\,\Delta, \mu\,\sigma_v)$ for any $\mu > 0$. For this reason, we use the one-variable function $K_{\text{opt}}(\chi)$ to refer to the optimal period from (2.25) for a particular χ.

In the context of the sawtooth K-periodic inputs (2.26), strategies that select K so as to keep a fixed sawtooth spacing δ_w achieve the minimum possible growth rate. In particular, in App. A.2 we show that, for any given χ, if we select the period K in (2.26) according to

$$K = \lceil \lambda\,\chi + 1 \rceil \tag{2.27}$$

where λ can be any positive constant, the associated worst-case information loss grows linearly with χ. In general, there is an optimal λ for any particular noise PDF $p_{\tilde{v}}(\cdot)$, resulting in an optimal normalized sawtooth spacing. Specifically, consider the normalized spacing between successive samples of $w[n]$ in (2.26), namely,

$$d(\chi; K) \triangleq \frac{\delta_w}{\sigma_v} = \frac{2\,\chi}{K-1}\,. \tag{2.28}$$

In addition, let $d_{\text{opt}}(\chi)$ denote the normalized spacing associated with the optimal period $K_{\text{opt}}(\chi)$ from (2.25), *i.e.*,

$$d_{\text{opt}}(\chi) \triangleq d(\chi; K_{\text{opt}}(\chi))\,. \tag{2.29}$$

In App. A.2, we outline a method for finding the asymptotic optimal normalized spacing

$$d_\infty \triangleq \lim_{\chi\to\infty} d_{\text{opt}}(\chi), \tag{2.30}$$

associated with a particular sensor noise PDF. For purposes of illustration, we also show in

App. A.2 that in the special case that the sensor noise is Gaussian with variance σ_v^2

$$d_\infty \approx 2.5851\,, \tag{2.31}$$

while the associated worst-case information loss is well approximated by

$$\mathcal{L}_{\max}^{\text{per}}(\chi) \approx 1.4754 \left(\frac{2\chi}{d_\infty} + 1\right) \tag{2.32}$$

for large χ. In this Gaussian scenario, if we select $w[n]$ as in (2.26) with $K = \lceil 2\chi/d_\infty + 1 \rceil$, the worst-case information loss is given by (2.32) and achieves the optimal growth rate for known control waveforms. We next extend the above analysis to quantizers with $M > 2$.

Optimized Periodic Waveforms for Signal Quantizers with $M > 2$

As we have seen in the preceding section, selection of $w[n]$ according to (2.26) for $M = 2$ results in a two-level quantizer with periodically time-varying thresholds uniformly spaced in $[-\Delta, \Delta]$. This selection method minimizes the maximum distance between the parameter value and the closest of the time-varying thresholds, over the dynamic range $(-\Delta, \Delta)$. The same strategy can be used for $M > 2$, although the availability of multiple thresholds allows for reduction of the dynamic range that $w[n]$ needs to span. We assume that all quantizer thresholds are within the dynamic range, *i.e.*, $-\Delta < X_i < \Delta$, for $i = 1, 2, \cdots, M-1$. In this case, the effective dynamic range Δ_{eff} that $w[n]$ needs to span is given by

$$\Delta_{\text{eff}} = \max_i \, \delta x_i\,,$$

where

$$\delta x_i = \begin{cases} X_1 + \Delta & \text{if } i = 1 \\ X_i - X_{i-1} & \text{if } 2 \le i \le M-2 \\ \Delta - X_{M-1} & \text{if } i = M-1 \end{cases}\,.$$

In particular, we consider using the control input (2.26) where the effective spacing between thresholds δ_w is given in terms of Δ and the quantizer thresholds $X_1, X_2, \cdots, X_{M-1}$, as

follows

$$\delta_w = \max_i \ \delta w_i \tag{2.33a}$$

where

$$\delta w_i = \begin{cases} \dfrac{2\,\delta x_i}{K-1} & \text{if } i = 1,\, M-1 \\ \dfrac{2\,\delta x_i}{K} & \text{if } 2 \le i \le M-2 \end{cases} . \tag{2.33b}$$

For any A in $(-\Delta,\, \Delta)$, this selection guarantees that at least one of the $M-1$ time-varying quantizer thresholds is within $\delta_w/2$ of the parameter, where δ_w is given by (2.33a). One can in principle perform the optimization (2.25) to obtain $K_{\text{opt}}(\Delta,\, \sigma_v)$ for any $F(\cdot)$ with $M > 2$. We should emphasize, however, that at high SNR we may often obtain an approximate estimate of performance via our results for the case $M = 2$. For instance, for $\Delta_{\text{eff}}/\sigma_v$ large and small enough λ in (2.27), the optimal normalized spacing and the corresponding worst-case information loss for a quantizer with $M > 2$ are approximately given by the respective quantities for the symmetric two-level quantizer, with χ replaced by $\chi_{\text{eff}} = \Delta_{\text{eff}}/\sigma_v$.

If in addition there is freedom in selecting the $M-1$ quantizer thresholds, these can be selected so that $\delta w_i = \delta w_j$ for all i and j in (2.33b) which implies that $\delta_w = \Delta/[(M-1)\,K-1]$. This selection guarantees that for every K successive observations, the collection of all $M\,K$ associated quantizer thresholds form a uniformly spaced collection in $[-\Delta,\, \Delta]$. For instance, in the special case that the sensor noise is Gaussian, the optimal normalized spacing and the worst-case loss for large χ are given by (2.31) and (2.32), respectively, with $\chi/(M-1)$ replacing χ on the left hand side of (2.32). In summary, simply constructed classes of periodic control waveforms achieve the optimal information loss growth rate with peak SNR.

2.2.3 Control Inputs in the Presence of Feedback

In this section we consider the scenario where, in addition to knowing the control waveform, the estimator has the option of using feedback from past output observations in the selection of future control input values. Specifically, we develop performance bounds for the problem

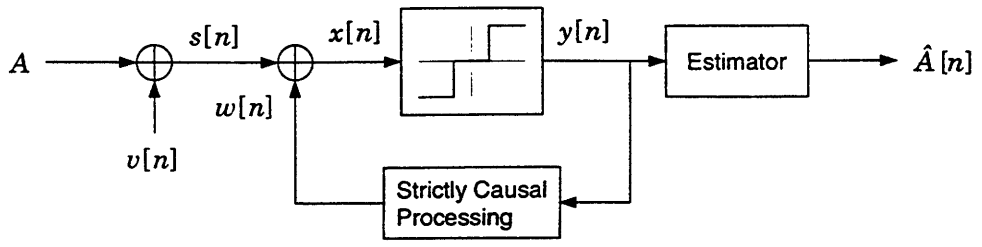


Figure 2-5: Estimation based on observations from a signal quantizer, where feedback from the quantized output is used in the selection of the control input.

of estimation of A based on $\mathbf{y}^N$, where the control input sequence $w[n]$ is a function of all past quantized observations. This scenario is depicted in Fig. 2-5 where $w[n] = g(\mathbf{y}^{n-1})$.

We next show that the worst-case information loss for any feedback-based control input strategy is lower bounded by the *minimum* possible information loss for the same quantizer system with $w[n] = 0$; in Section 2.3 we develop feedback-based control selection algorithms that effectively achieve this lower bound. Examination of the Cramér-Rao bound (2.23) reveals that for any A in $(-\Delta, \Delta)$ we can obtain information loss equal to $\mathcal{L}(A_o)$ by selecting $w[n] = A_o - A$. In particular, if there exists a parameter value A_* for which $\mathcal{B}(A;\ y) \geq \mathcal{B}(A_*;\ y)$ for all A in $(-\infty, \infty)$ and where $\mathcal{B}(A;\ y)$ is given by (2.12) with α replaced by v, then using (2.23) we obtain

$$\mathcal{B}(A;\ \mathbf{y}^N,\ \mathbf{w}^N) \geq \mathcal{B}(A_*;\ y)/N\,, \tag{2.34}$$

with equality achieved for $w[n] = A_* - A$ for $n = 1, 2, \cdots, N$. This control input results in

$$\mathcal{L}(A; \mathbf{w}^N) \geq \mathcal{L}(A; A_* - A) = \mathcal{L}(A_*)\,, \tag{2.35}$$

where $\mathcal{L}(A)$ is given by (2.4), and where $\mathcal{B}(A;\ y)$ is given by (2.12) with α replaced by v.

The minimum information loss from (2.35) decreases as the number of quantization levels increases. In App. A.3 we show that as we would expect, the minimum information loss $\mathcal{L}(A_*)$ tends to zero as the number of quantization levels approaches infinity for any sensor noise.

For a number of common sensor noises the control-free information loss for the system corresponding to $M = 2$ is minimized at the negative of the median of the PDF $p_v(\cdot)$, *i.e.*, $C_v(-A_*) = 1/2$. The corresponding minimum information loss (2.35) can be obtained by

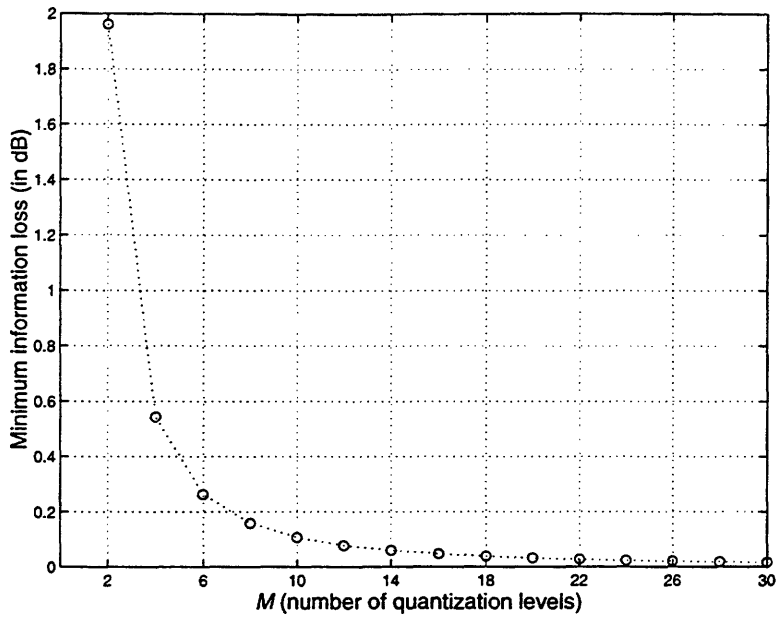


Figure 2-6: Minimum possible information loss as a function of quantization levels for a uniform quantizer in IID Gaussian noise. For any given M, the threshold spacing is selected so as to minimize this loss.

evaluating (2.4) at $A = A_*$, while employing (2.8) and (2.14) for $\sigma_w = 0$, namely,

$$\mathcal{L}(A_*) = \left[4\, p_{\tilde{v}}^2 \left(-A_*/\sigma_v\right)\, \mathcal{B}\left(0;\, \tilde{s}\right)\right]^{-1} \ , \tag{2.36}$$

which is actually independent of σ_v and Δ, since $-A_*/\sigma_v$ equals the median of the PDF of $\tilde{v}[n]$.

Special Case: Gaussian Sensor Noise

In the case that the sensor noise is Gaussian, the minimum information loss (2.35) decays rapidly to zero as more quantization levels are introduced. In Fig. 2-6 we plot the minimum possible information loss through any uniform M-level quantizer for various values of M, in the presence of IID Gaussian noise. From the figure it is apparent that a few quantization levels suffice to effectively eliminate the minimum information loss due to quantizer-based processing.

For the two-level quantizer (2.13) in this Gaussian scenario, use of (2.16) for $\sigma_\alpha = \sigma_v$ in

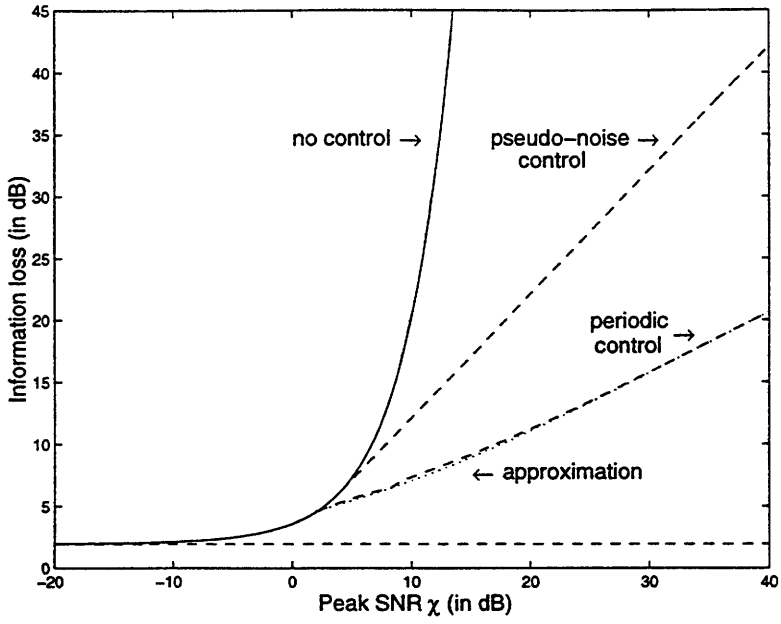


Figure 2-7: Worst-Case information loss over $|A| < \Delta$ for a two-level quantizer in zero-mean IID Gaussian noise of variance σ_v^2, with no control input (solid), pseudo-noise control inputs (upper dashed), and known periodic control waveforms (middle dashed). The dotted curve depicts approximation (2.32). The lower dashed line depicts the minimum possible information loss (≈ 2 dB) for any control input scheme.

(2.7) reveals that $A_* = 0$. In this case, (2.34) reduces to

$$\mathcal{B}\left(A;\ \mathbf{w}^N,\ \mathbf{y}^N\right) \geq \mathcal{B}\left(0;\ y\right)/N = \frac{\pi\,\sigma_v^2}{2\,N}\,, \tag{2.37}$$

while from (2.36) the information loss for any parameter value A is lower-bounded as follows

$$\mathcal{L}\left(A; \mathbf{w}^N\right) \geq \mathcal{L}(0) = \frac{\pi}{2}\,, \tag{2.38}$$

which corresponds to a 2 dB information loss.

Fig. 2-7 depicts the worst-case information loss for the system corresponding to $M = 2$ in the context of Gaussian sensor noise and the various control input scenarios that we have examined. As the figure reflects, the performance of the control-free system (solid curve) degrades rapidly as the peak SNR is increased. The benefits of pseudo-noise control inputs (upper dashed curve) at high peak SNR are clearly evident, and known periodic control inputs provide additional performance benefits (middle dashed curve) over pseudo-noise control inputs. In particular, the associated worst-case information loss increases linearly

with peak SNR as the accurate approximation (2.32) reveals. Finally, in the presence of feedback from quantized output to the control input, the performance is lower bounded by the minimum possible information loss of 2 dB, which is independent of χ. In Section 2.3 we develop control selection strategies and associated estimators that meet all these bounds.

2.3 Efficient Estimation

In this section we develop control input selection strategies and associated estimators which achieve the performance limits computed in Section 2.2. A natural measure of performance of a specific system comprising a control input a quantizer and a particular estimator is the *MSE loss*, which we define as the ratio of the actual MSE of a particular estimator of A based on observation of $\mathbf{y}^N$, divided by the Cramér-Rao bound for estimating A from observation of $\mathbf{s}^N$. In case an efficient estimator of A based on $\mathbf{s}^N$ exists, the notion of the MSE loss of any given estimator of A given $\mathbf{y}^N$ has an alternative, appealing interpretation. In this case, the MSE loss represents the additional MSE in dB that arises from estimating A using this particular estimator on $\mathbf{y}^N$, instead of efficiently estimating A via $\mathbf{s}^N$. Analogously to $\mathcal{L}_{\max}$ in (2.7), the worst-case MSE loss of an estimator is defined as the supremum of the MSE loss function over the range $|A| < \Delta$.

In this section we construct estimators for which the corresponding MSE loss asymptotically achieves the associated information loss, for each of the control input scenarios of Sec 2.2. We examine the control-free and pseudo-noise control scenarios first, and then develop estimators applicable to known K-periodic control inputs. Finally, in the context of feedback we develop control input selection strategies and associated estimators which achieve the minimum possible information loss for any given system.

2.3.1 Pseudo-noise Control Inputs

For pseudo-noise control inputs, the maximum-likelihood (ML) estimator of A based on $\mathbf{y}^N$ over the restricted dynamic range $|A| \leq \Delta$ satisfies

$$\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N; \Delta\right) = \underset{|\theta| \leq \Delta}{\arg\max}\ \ln P\left(\mathbf{y}^N; \theta\right) \ , \tag{2.39}$$

where $\ln P\left(\mathbf{y}^N;\theta\right)$ is the log-likelihood function given by (2.9). We first examine ML estimation for the system with $M = 2$, and then construct estimators for signal quantizers with $M > 2$. Estimators of A for control-free systems can be readily obtained as a special case of the estimators of A for the associated systems with pseudo-noise control inputs by setting $\sigma_w = 0$.

ML Estimation for Signal Quantizers with $M = 2$ in IID Noise

If $F(\cdot)$ is given by (2.13) and $\alpha[n]$ is admissible, the ML estimator (2.39) can be found in closed form, by setting to zero the partial derivative of the log-likelihood function (2.9) with respect to A, *viz.*,

$$\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right) = \mathcal{I}_\Delta\left(\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\infty\right)\right) \tag{2.40}$$

where $\mathcal{I}_\Delta(\cdot)$ is the following piecewise-linear limiter function

$$\mathcal{I}_\Delta(x) = \begin{cases} x & \text{if } |x| \le \Delta \\ \Delta\ \mathrm{sgn}(x) & \text{otherwise} \end{cases} . \tag{2.41}$$

The function $\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\infty\right)$ denotes the ML estimate of A from $\mathbf{y}^N$ when there are no restrictions imposed in the dynamic range of the unknown parameter A.[4] In particular,

$$\begin{aligned} \hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\infty\right) &= \arg\max_{\theta}\ \ln P\left(\mathbf{y}^N;\theta\right) \\ &= -C_\alpha^{-1}\left(\frac{\mathcal{K}_1\left(\mathbf{y}^N\right)}{N}\right) , \end{aligned} \tag{2.42}$$

where $C_\alpha^{-1}(\cdot)$ in (2.42) is the inverse of $C_\alpha(\cdot)$, and $\mathcal{K}_{Y_i}\left(\mathbf{y}^N\right)$ denotes the number of elements in $\mathbf{y}^N$ that are equal to Y_i. In the special case that $w[n]$ and $v[n]$ are zero-mean IID Gaussian noise sequences with variances σ_w^2 and σ_v^2, respectively, (2.42) reduces to

$$\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\infty\right) = -\sigma_\alpha\, Q^{-1}\left(\frac{\mathcal{K}_1\left(\mathbf{y}^N\right)}{N}\right) . \tag{2.43}$$

For any parameter value A in the range $(-\Delta, \Delta)$, the Cramér-Rao bound (2.14) is a

[4] Note that (2.40) does not necessarily hold for $M > 2$.

reasonable predictor of the MSE performance of the ML estimator (2.40)–(2.42) provided that the number of observations N is large enough. Indeed, as shown in App. A.4 for any $A \in (-\Delta, \Delta)$, the ML estimator (2.40)–(2.42) is asymptotically efficient in the sense that it achieves the Cramér-Rao bound for unbiased estimates (2.14) for large enough N, *i.e.*,

$$\lim_{N\to\infty} N E\left[\left(\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right) - A\right)^2\right] = \mathcal{B}\left(A;\ y\right) .$$

Although the ML estimate (2.40)–(2.42) is asymptotically unbiased and efficient for any A in $(-\Delta, \Delta)$, the associated MSE does not converge uniformly to the Cramér-Rao bound in the parameter A with N. Specifically, for any fixed N, no matter how large, there exist parameter values close enough to the boundaries $\pm\Delta$ for which the ML estimator has significant bias,[5] in which case (2.14) should not be expected to accurately predict the associated MSE of the ML estimator. This is clearly reflected in Fig. 2-8, where the actual MSE loss for $\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right)$ is also depicted alongside the associated information loss for the Gaussian noise scenario. In particular, the dashed and solid lines depict the MSE loss from Monte-Carlo simulations for the ML estimator (2.40)–(2.42), in the absence ($\sigma_w = 0$) and presence ($\sigma_w = 2/\pi$) of pseudo-noise control input, respectively, for $\sigma_v = 0.1$, $\Delta = 1$, and $N = 100,\ 10^4$. As we can see in Fig. 2-8, when the pseudo-noise level is $\sigma_w = 2/\pi$ the worst-case MSE loss is about 21 dB. However, in the absence of a control input, the worst-case MSE loss is about 36 dB for $N = 100$, and 55 dB for $N = 10^4$. For both values of N the Cramér-Rao bound (2.14) is applicable for only a subset of the dynamic range, whose size increases with N. In fact, since the ML estimator is asymptotically efficient for any $|A| < \Delta$ with respect to the Cramér-Rao bound (2.14) for unbiased estimates, the worst-case MSE loss for the control-free system increases with N towards the associated worst-case information loss (2.20), which is approximately 211 dB.

ML Estimation for Signal Quantizers with $M > 2$ in IID Gaussian Noise

For the estimation problem (2.1)–(2.2) where $F(\cdot)$ is an M-level quantizer and $\alpha[n]$ is an IID sequence, the set of sufficient statistics reduces to $\mathcal{K}_{Y_1}\left(\mathbf{y}^N\right), \cdots, \mathcal{K}_{Y_{M-1}}\left(\mathbf{y}^N\right)$ (*cf.* (2.9)).

[5] By incorporating the bias of the ML estimator (2.40)–(2.42) it is possible to obtain a Cramér-Rao bound that directly applies to the associated MSE. An even tighter bound can be obtained by properly combining three separate Cramér-Rao bounds, each describing the effects of a piecewise linear region of the soft limiter $\mathcal{I}_\Delta(\cdot)$ on $\hat{A}_{\mathrm{ML}}\left(A;\infty\right)$ in (2.40).

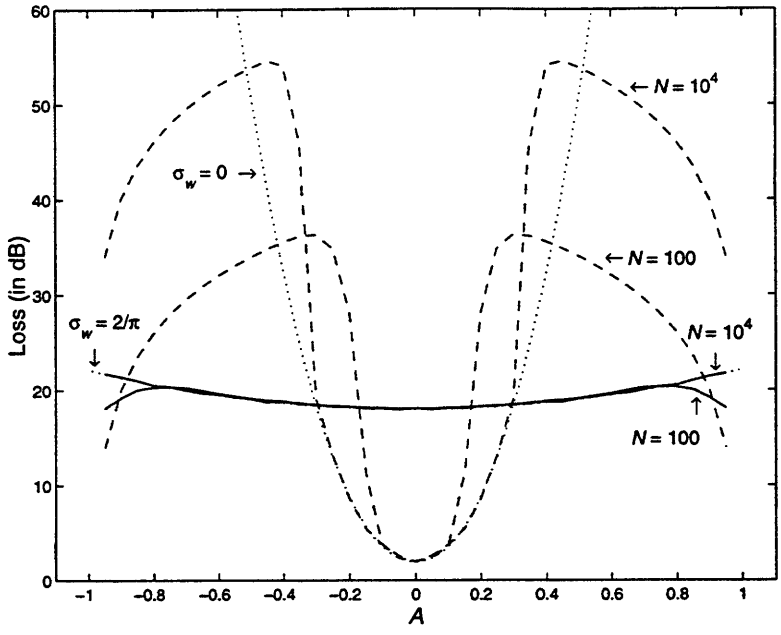


Figure 2-8: MSE loss from Monte-Carlo simulations for a system comprising a Gaussian pseudo-noise control input a two-level quantizer and the ML estimator (2.40)–(2.42) for $\Delta = 1$, $\sigma_v = 0.1$ and various pseudo-noise power levels. The dashed curves depict the MSE loss of $\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right)$ in the absence of control input (*i.e.*, $\sigma_w = 0$); upper curve: $N = 10^4$, lower curve: $N = 100$. The solid curves depict the MSE loss of $\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right)$ for $\sigma_w = 2/\pi$, and for $N = 100,\ 10^4$. For comparison, the associated information loss functions are depicted by the dotted curves (also shown in Fig. 2-3).

For the special case that $\alpha[n]$ is Gaussian with variance σ_α^2, we develop in App. A.5 an EM algorithm [14] for obtaining the ML estimate (2.39). This algorithm takes the following form:

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_\Delta\left(\hat{A}_{\mathrm{EM}}^{(k)} + \frac{\sigma_\alpha}{\sqrt{2\,\pi}\,N}\sum_{m=1}^{M}\mathcal{K}_{Y_m}(\mathbf{y}^N)\,\frac{\exp\left(-\frac{\left(X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}\right)^2}{2\,\sigma_\alpha^2}\right)-\exp\left(-\frac{\left(X_m-\hat{A}_{\mathrm{EM}}^{(k)}\right)^2}{2\,\sigma_\alpha^2}\right)}{Q\left(\frac{X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}}{\sigma_\alpha}\right)-Q\left(\frac{X_m-\hat{A}_{\mathrm{EM}}^{(k)}}{\sigma_\alpha}\right)}\right), \tag{2.44}$$

initialized with $\hat{A}_{\mathrm{EM}}^{(0)} = 0$. Provided that the log-likelihood function does not possess multiple local minima, (2.44) provides the ML estimate (2.39), *i.e.*,

$$\hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N;\Delta\right) = \lim_{k\to\infty}\hat{A}_{\mathrm{EM}}^{(k)}\,.$$

Empirical evidence suggests that $\lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}$ obtained via the algorithm (2.44) is asymptotically efficient, *i.e.*, it achieves (2.12) for large N. Consequently, use of information loss as an accurate predictor of the MSE loss is also justified in this scenario.

Efficient Estimation for Signal Quantizers with $M > 2$ in IID Noise

In general, there is no computationally efficient method for obtaining the ML estimate (2.39) of A in nonGaussian noise via a signal quantizer with $M > 2$. In this section we present an alternative class of elementary estimators which can be shown to be asymptotically efficient for any admissible noise PDF $p_\alpha(\cdot)$, in the sense that for any $|A| < \Delta$ the MSE of the estimator approaches the bound (2.12) for large N.

Without loss of generality we may view the output of the quantizer $F(\cdot)$ in (2.2) as the collection of the outputs of $M-1$ two-level quantizers generating the following observed sequences

$$y_i[n] = \mathrm{sgn}\ (x[n] - X_i) \qquad i = 1, 2, \cdots, M-1,$$

where $x[n] = s[n] + \alpha[n]$ (*cf.* Fig. 2-2) and the X_i's are the thresholds of the quantizer. Consider the ML estimates of A formed from each of these binary sequences, namely,

$$\hat{A}_i = \mathcal{I}_\Delta\left(\hat{A}_{\mathrm{ML}}\left(\mathbf{y}_i^N; \infty\right) + X_i\right) \qquad i = 1, 2, \cdots, M-1, \tag{2.45}$$

where

$$\mathbf{y}_i^N \triangleq \begin{bmatrix} y_i[1] & y_i[2] & \cdots & y_i[N] \end{bmatrix}^T,$$

and where $\mathcal{I}_\Delta(\cdot)$ is given by (2.41), and $\hat{A}_{\mathrm{ML}}(\cdot;\infty)$ is given by (2.42) with α replaced by v. In App. A.6 we show that the joint cumulative distribution of

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 & \cdots & \hat{A}_{M-1} \end{bmatrix}^T \tag{2.46}$$

approaches the cumulative distribution of a Gaussian random vector with mean $A\,\mathbf{1}$ (where $\mathbf{1}$ denotes a vector of 1's) and covariance matrix C/N, whose inverse is given by (A.39). We

also show in the appendix that if we use

$$\hat{A} = (\mathbf{1}^T \hat{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \hat{C}^{-1} \hat{\mathbf{A}} \tag{2.47}$$

where $\hat{C} = C(\hat{A}_i)$ for some $1 \leq i \leq M-1$, the estimator $\hat{A}$ is asymptotically efficient, *i.e.*,

$$\lim_{N \to \infty} N E\left[(\hat{A} - A)^2; A\right] = \mathcal{B}(A;\ y)\ , \tag{2.48}$$

where $\mathcal{B}(A;\ y)$ is given by (2.12). In practice, in computing $\hat{C}$ we may select the value of i for which $\mathcal{B}\left(\hat{A}_i;\ \mathbf{y}_i^N\right)$ is minimum, so as to expedite the MSE convergence to the asymptotic performance predicted by (2.48). In summary, the estimator first obtains the set (2.46) by means of (2.45) and (2.41)–(2.42), it then selects the value of i for which $\mathcal{B}\left(\hat{A}_i;\ \mathbf{y}_i^N\right)$ is minimized and forms $\hat{C} = C(\hat{A}_i)$, and finally substitutes $\hat{A}_i$ and $\hat{C}$ in (2.47) to obtain the asymptotically efficient estimate $\hat{A}$.

2.3.2 Known Control Inputs

In this section we construct estimators that exploit detailed knowledge of the applied control waveform. In particular, in the context of K-periodic control inputs that are known for estimation, we develop estimators that are asymptotically efficient in the sense that they asymptotically achieve (2.23).

For IID Gaussian sensor noise, the ML estimate of A from $\mathbf{y}^N$ given a control vector $\mathbf{w}^N$, where $w[n]$ is a K-periodic sequence and N is a multiple of K, can be obtained as a special case of the EM algorithm presented in App. A.5. In particular, the EM algorithm takes the following form

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)} + \sum_{\substack{1 \leq \ell \leq K \\ 1 \leq m \leq M}} \frac{\sigma_v \mathcal{K}_{Y_m}(\mathbf{y}^{\bar{N}}[\ell])}{\sqrt{2\pi N}} \frac{\exp\left(-\frac{\left(X_{m-1} - \hat{A}_{\mathrm{EM}}^{(k)} - w[\ell]\right)^2}{2\sigma_v^2}\right) - \exp\left(-\frac{\left(X_m - \hat{A}_{\mathrm{EM}}^{(k)} - w[\ell]\right)^2}{2\sigma_v^2}\right)}{Q\left(\frac{X_{m-1} - \hat{A}_{\mathrm{EM}}^{(k)} - w[\ell]}{\sigma_v}\right) - Q\left(\frac{X_m - \hat{A}_{\mathrm{EM}}^{(k)} - w[\ell]}{\sigma_v}\right)}\right), \tag{2.49a}$$

and

$$\hat{A}_{\mathrm{ML}} = \lim_{k \to \infty} \hat{A}_{\mathrm{EM}}^{(k)}\ , \tag{2.49b}$$

where $\bar{N} = N/K$, and $\mathbf{y}^{\bar{N}}[\ell]$ is the $\bar{N} \times 1$ vector comprised of the elements of the ℓth K-decimated subsequence, *i.e.*,

$$\mathbf{y}^{\bar{N}}[\ell] \stackrel{\Delta}{=} \begin{bmatrix} y[\ell] & y[K+\ell] & \cdots & y[N-K+\ell] \end{bmatrix}^T \qquad \ell = 1, 2, \cdots, K \ . \tag{2.50}$$

Empirical evidence suggests that the estimate resulting from the EM algorithm (2.49) is asymptotically efficient, *i.e.*, it achieves the Cramér-Rao bound (2.24) for large enough N.

Asymptotically efficient estimators in the context of nonGaussian sensor noises can be obtained in a fashion similar to those developed in App. A.6. Specifically, in the case $M = 2$, we may consider the vector $\hat{\mathbf{A}}$ in (2.46) where we use for $\hat{A}_i$ the ML estimate of A given the ith K-decimated subsequence from (2.50), *i.e.*,

$$\hat{A}_i = \mathcal{I}_\Delta \left(\hat{A}_{\mathrm{ML}} \left(\mathbf{y}^{\bar{N}}[i]; \infty \right) - w[i] \right) \qquad i = 1, 2, \cdots, K \tag{2.51}$$

and where $\mathcal{I}_\Delta(\cdot)$ and $\hat{A}_{\mathrm{ML}}(\cdot; \infty)$ are given by (2.41) and (2.42), respectively. The $\hat{A}_i$'s from (2.51) are independent random variables, since for any $i \neq j$, $\mathbf{y}^{\bar{N}}[i]$ and $\mathbf{y}^{\bar{N}}[j]$ are independent random vectors. Therefore, the corresponding vector $\hat{\mathbf{A}}$ from (2.46) is asymptotically Gaussian (in terms of its cumulative distribution), with diagonal covariance matrix C/N; the (i, i)th entry of the matrix C equals $\mathcal{B}(A + w[i];\ y[i])$, where $\mathcal{B}(A;\ y)$ is given by (2.12) with α replaced by v. Consequently, an asymptotically efficient estimate is provided by $\hat{A}$ from (2.47); the estimate covariance matrix that is used for faster MSE convergence to the asymptotic performance is given by $\hat{C} = C(\hat{A}_i)$ where i is the index that minimizes $\mathcal{B}\left(\hat{A}_i + w[i];\ \mathbf{y}^{\bar{N}}[i]\right)$.

Asymptotically efficient estimators can also be constructed for signal quantizers with $M > 2$ and known K-periodic inputs in nonGaussian sensor noise. Specifically, for each M-ary subsequence $\mathbf{y}^{\bar{N}}[\ell]$ from (2.50) we may first apply the algorithm (2.45)–(2.47) to obtain K statistically independent estimates of A. By combining these K estimates in a fashion similar to the method used in the case $M = 2$ for combining the estimates (2.51), we obtain an asymptotically efficient estimator of A based on $\mathbf{y}^N$ given $\mathbf{w}^N$.

2.3.3 Control Inputs in the Presence of Feedback

In Section 2.2.3 we have shown that the worst-case information loss of a system composed of a signal quantizer and an additive control input is lower-bounded by the minimum possible

information loss of the same system in the control-free case. In this section we develop control input selection strategies based on past quantized output samples and construct associated estimators which effectively achieve this bound.

Feedback Control and Estimation for Signal Quantizers with $M = 2$

We first examine the Gaussian sensor noise scenario with $M = 2$ in detail. As (2.38) reveals, the associated control-free information loss is minimized for $w[n] = -A$. Although this control input selection is not permissible, it suggests a viable control input selection method based on past quantized observations. Specifically, if $\hat{A}[n]$ is any consistent estimator of A based on $\mathbf{y}^n$, a reasonable choice for the control input sequence is as follows

$$w[n] = -\hat{A}[n-1] \; . \tag{2.52}$$

Assuming the control sequence is selected according to (2.52), the ML estimator at time n satisfies

$$\hat{A}_{\mathrm{ML}}[n] = \underset{|\theta| \leq \Delta}{\arg\max} \sum_{m=1}^{n} \ln Q\left(y[m] \; \left(\hat{A}_{\mathrm{ML}}[m-1] - \theta\right)\right) \; .$$

In App. A.5 we show that in the Gaussian scenario the ML estimate of A based on $\mathbf{y}^n$ for $n = 1, 2, \cdots$ can be obtained using the following EM algorithm,

$$\hat{A}_{\mathrm{EM}}^{(k+1)}[n] = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)}[n] + \frac{\sigma_v}{\sqrt{2\pi n}} \sum_{m=1}^{n} y[m] \, \frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{ML}}[m-1] - \hat{A}_{\mathrm{EM}}^{(k)}[n]\right)^2}{2\sigma_v^2}\right)}{Q\left(y[m] \, \frac{\hat{A}_{\mathrm{ML}}[m-1] - \hat{A}_{\mathrm{EM}}^{(k)}[n]}{\sigma_v}\right)}\right) , \tag{2.53a}$$

initialized with $\hat{A}_{\mathrm{EM}}^{(0)}[n] = \hat{A}_{\mathrm{ML}}[n-1]$ and $\hat{A}_{\mathrm{ML}}[0] = 0$, where for any n,

$$\hat{A}_{\mathrm{ML}}[n] = \lim_{k \to \infty} \hat{A}_{\mathrm{EM}}^{(k)}[n] \; . \tag{2.53b}$$

Although empirical evidence suggests that the ML estimator obtained by means of the EM algorithm in (2.53) achieves the 2 dB information loss bound (2.38) for any A in $(-\Delta, \Delta)$ for a moderate number of observations,[6] it is rather computationally intensive; for any

[6]There are a number of other control input selection methods and associated estimators which can

additional observed sample an EM algorithm has to be employed. In addition, even though the number of iterations necessary for adequate convergence of the EM algorithm appears to be small for large n, the algorithm may still be impractical.

We next develop algorithms that achieve the bound (2.38) and have the additional advantage that they can be implemented very efficiently. These are based on the observation that once the estimate $\hat{A}[n]$ is not changing significantly with n (*i.e.*, the changes are small with respect to σ_v) we may assume that $A+w[n+1]$ is in the regime where the information loss is small, and a linear estimator can be used that approaches the 2 dB bound (2.38). Specifically, let $z = Q\left(A/\sigma_v\right)$ and assume that $|A/\sigma_v| < 0.1$. In this regime, the truncated power series expansion provides a reasonable approximation for $Q^{-1}\left(z\right)$, *i.e.*,

$$Q^{-1}\left(z\right) \approx \sqrt{\frac{\pi}{2}}(1-2z)\,. \tag{2.54}$$

We can use (2.54) to form a linear estimator as follows. Assuming that the estimation error is inversely proportional to the measurements (which implies that the asymptotic MSE loss is not infinite), the estimate at time n is given as a weighted sum of the estimate at time $n-1$ and an estimate arising from using the nth measurement $y[n]$ alone, *i.e.*,

$$\hat{A}_{\mathrm{L}}[n] = \frac{n-1}{n}\,\hat{A}_{\mathrm{L}}[n] + \frac{1}{n}\,\hat{A}[n|y[n]]\,, \tag{2.55}$$

where the estimate based on the nth measurement alone is given by using (2.54) in (2.43) (by setting σ_w to 0), and the fact that $w[n] = -\hat{A}_{\mathrm{L}}[n-1]$, *i.e.*,

$$\hat{A}[n|y[n]] = \hat{A}_{\mathrm{L}}[n-1] + \sigma_v\sqrt{\frac{\pi}{2}}\,y[n]\,. \tag{2.56}$$

By incorporating (2.56) in (2.55) this linear estimator takes the following iterative form

$$\hat{A}_{\mathrm{L}}[n] = \hat{A}_{\mathrm{L}}[n-1] + \sigma_v\sqrt{\frac{\pi}{2}}\,\frac{y[n]}{n}\,. \tag{2.57}$$

In order to obtain an algorithm that converges much faster than (2.57) to the 2 dB bound

approach arbitrarily close to the 2 dB bound; the systems developed in this chapter for the case $M > 2$ and nonGaussian noise are such an example. However, the associated MSE of these algorithms converges to the bound (2.38) considerably slower than the algorithms of this section. In fact, the number of samples required so that the MSE of (2.53) with $w[n]$ as in (2.52) effectively achieves the 2 dB bound (2.38) increases linearly with $\ln(\chi)$.

(2.38), we employ the EM algorithm (2.53) for $n \leq n_o$ and the recursive algorithm (2.57) for $n > n_o$, *i.e.*,

$$\check{A}[n] = \begin{cases} \hat{A}_{\mathrm{ML}}[n] \quad \text{from (2.53)} & \text{if } n \leq n_o \\ \mathcal{I}_{\Delta}\left(\check{A}[n-1] + \sigma_v \sqrt{\frac{\pi}{2}} \frac{y[n]}{n}\right) & \text{if } n > n_o \end{cases}, \tag{2.58}$$

where the control input $w[n]$ is given by (2.52) provided that we substitute $\check{A}[n-1]$ for $\hat{A}[n-1]$, and where we also incorporated the dynamic range information by means of $\mathcal{I}_{\Delta}(\cdot)$.

Selection of an appropriate value for n_o is related to the peak SNR χ. Since, in principle, the larger the peak SNR, the longer (in terms of the number of observations) it takes $A - \hat{A}_{\mathrm{ML}}[n]$ to reach the linear regime (2.54), we consider the case $\Delta \gg \sigma_v$. For instance, assume we are interested in selecting n_o so that the $\sqrt{\mathrm{MSE}}$ in $\check{A}[n_o]$ is less than a given fraction of σ_v (so that the truncated series approximation is valid), for example $\sigma_v/8$. For small enough n_o, the maximum MSE from n_o observations is roughly given as the square of $\Delta\, 2^{-n_o}$. In summary, this crude-MSE based rule of thumb for selecting n_o reduces to $n_o \geq \log_2(\Delta/\sigma_v) + 3$.

The solid and dashed curves in Fig. 2-9 depict the MSE of the ML estimator obtained by means of the EM algorithm in (2.53), and of the computationally efficient estimator (2.58) with $n_o = 10$, respectively, based on Monte-Carlo simulations. The system parameters for this simulation are $\Delta = 1$, $\sigma_v = 0.1$, resulting in $\log_2(\Delta/\sigma_v) \approx 6.6$, while $A = 0.4$. In both cases the control sequence is selected according to (2.52). The lower and upper dotted lines depict $\mathcal{B}\left(A;\, \mathbf{s}^N\right)$ and the right hand side of (2.37), respectively. As we can see in this figure, both estimates effectively achieve the 2 dB loss bound (2.38) for a moderate number of observations.

In terms of the actual implementation of the estimator (2.58), for a given n_o there are 2^{n_o} possible values of $\hat{A}_{\mathrm{ML}}[n_o]$. These 2^{n_o} estimate values can be precomputed and stored in a lookup table. This results in an appealing computationally efficient implementation, whereby given n_o or fewer observations the estimate is obtained from a lookup table, while once the number of observations exceeds n_o, a recursive linear estimator is employed. Since n_o grows logarithmically with χ, the number of lookup table entries for storing all possible values of $\hat{A}_{\mathrm{ML}}[n_o]$ grows only linearly with peak SNR χ.

A similar strategy can be used in the context of quantizer systems using feedback in any

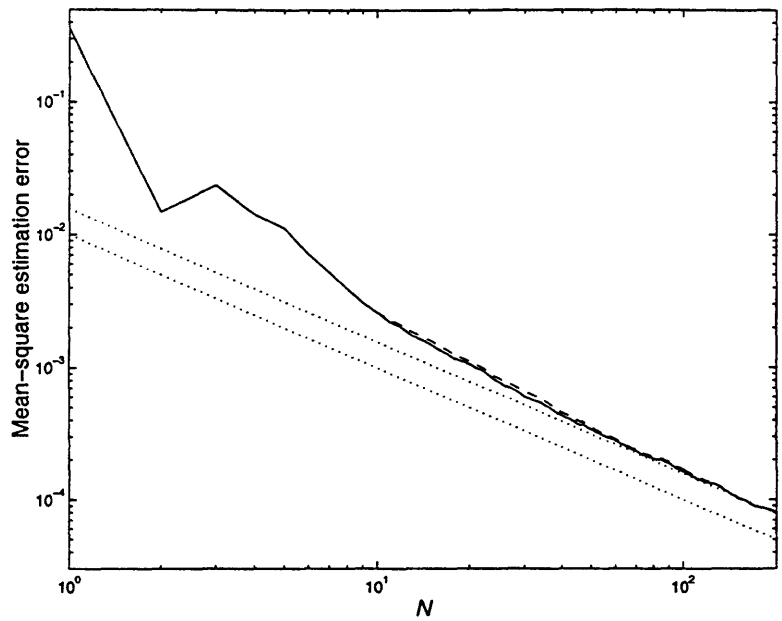


Figure 2-9: MSE from Monte-Carlo simulations for $\hat{A}_{\mathrm{ML}}[n]$ (solid) and $\check{A}[n]$ with $n_o = 10$ (dashed), based on observations from a signal quantizer with $M = 2$ exploiting feedback according to (2.52). The lower dotted line represents the Cramér-Rao bound for estimating A based on $s[n]$, while the upper dotted line is the 2 dB bound (2.38); Parameters: $\sigma_v = 0.1$, $\Delta = 1$, and $A = 0.4$.

sensor noise. In the general case A_* in (2.35) may not equal zero. A reasonable extension of the control input selection method (2.52) for nonzero A_* is as follows

$$w[n] = A_* - \hat{A}[n-1]\,. \tag{2.59}$$

An estimator similar to (2.58) can be used to estimate A in this case. Specifically, for $n \leq n_o$ the estimator may consist of a precomputed lookup table, while for $n > n_o$ a recursive estimator resulting from a truncated series expansion of $C_v^{-1}(z)$ around $z = A_*$ can be employed, namely,

$$\check{A}[n] = \mathcal{I}_\Delta\left(\check{A}[n-1] + \frac{1}{n}\,\frac{y[n] + 1 - 2\,C_v(-A_*)}{2\,p_v(-A_*)}\right).$$

In particular, if A_* is the median of $p_v(\cdot)$, in which case $\mathcal{L}(A_*)$ is given by (2.36), we have

$$\check{A}[n] = \mathcal{I}_\Delta\left(\check{A}[n-1] + \frac{y[n]}{2\,n\,p_v(-A_*)}\right) \qquad \text{for } n > n_o\,.$$

In general, empirical evidence suggests that the MSE loss of these algorithms practically achieves the associated $\mathcal{L}(A_*)$ for a moderate number of observations.

Feedback Control and Estimation for Signal Quantizers with $M > 2$

For the Gaussian sensor noise scenario, the EM algorithm (2.53) can be extended to $F(\cdot)$ with $M > 2$; the resulting algorithm is a special case of the one presented in App. A.5. Empirical evidence suggests that it is also asymptotically efficient. Assuming flexibility in selecting the thresholds of the M-level quantizer, the corresponding information loss (2.35) can be obtained from Fig. 2-6. For instance, for the optimal selection of the quantizer thresholds for $M = 6$ we have $A_* = 0$; if the control input is selected according to (2.59), the EM algorithm in App. A.5 yields a worst-case MSE loss of about 0.25 dB. Similarly to $\mathcal{L}_{\max}$, the asymptotic MSE loss is independent of σ_v and Δ.

For signal quantizers with $M > 2$ where $v[n]$ is any nonGaussian noise, we may use the following two stage approach that effectively achieves $\mathcal{L}(A_*)$. For the first N_1 observations we may employ any consistent estimator $\hat{A}_1[n]$ of A. For instance, we may use one of the feedback-based algorithms corresponding to the system $M = 2$ by ignoring all but two of the M levels of the quantized output. In the second stage, we fix $w[n] = A_* - \hat{A}_1[N_1]$ for all $n > N_1$. The number N_1 determines the accuracy of the approximation

$$\mathcal{L}\left(A_* + A - \hat{A}_1[N_1]\right) \approx \mathcal{L}(A_*) \; .$$

For any given $n > N_1$, we can then obtain an estimate $\hat{A}_2[n]$ of A from

$$\begin{bmatrix} y[N_1+1] & y[N_1+2] & \cdots & y[n] \end{bmatrix}^T \; ,$$

by means of (2.45)–(2.47), which is asymptotically efficient with respect to $\mathcal{L}\left(A_* + A - \hat{A}_1[N_1]\right)$. For faster convergence, the overall estimate can be a weighed sum of the estimates $\hat{A}_1[N_1]$ and $\hat{A}_2[n]$. Although the associated asymptotic MSE loss can be made to approach arbitrarily close to $\mathcal{L}(A_*)$, these algorithms typically require significantly larger data sets to effectively achieve the desired information loss, as compared to the algorithms for $M = 2$ of the previous section.

Chapter 3

Static Case Extensions for Quantizer Bias Control Systems

In a number of applications involving estimation of slowly-varying information-bearing signals, data may be collected from multiple sensors. In this case, the acquired measurements must be efficiently encoded at each sensor and, in turn, these encoded streams must be effectively combined at the host to obtain accurate signal estimates. In addition, irrespective of whether the application involves one or multiple sensors, a number of other issues may arise and may thus have to be taken into consideration. For instance, we often have available accurate signal models or other forms of prior information about the information-bearing signal. In such cases, we would be interested in exploiting any such form of additional information to improve the quality of the encodings and the associated estimates from these encodings. In addition, there are many instances where the noise at each sensor is non-stationary, or its statistical characterization is only partially known. It is important to incorporate such forms of uncertainty in the encoding and estimation algorithms and to determine the extent to which such issues may affect the overall system performance.

In this chapter we develop a number of such extensions of the systems we examined in Chapter 2. These extensions address a representative collection of cases of the signal estimation problem from digitally encoded measurements that may arise in practice. In Section 3.1 we consider estimation of a static signal from digitally encoded data obtained from multiple sensors and develop multi-sensor extensions of the signal encoding strategies and the algorithms of Chapter 2. As we show, the performance and optimal design of the

encoding and estimation algorithms of many multi-sensor generalizations of the estimation problem of Chapter 2 are natural extensions of the associate single-sensor performance and algorithms.

In Section 3.2 we consider the case where we have available *a priori* information about the relative likelihood of values of the information-bearing signal and would like to exploit it so as to improve the estimate quality. For instance, spatial or temporal correlations of the information-bearing signal are often available and can often be used to obtain such an *a priori* description of the static parameter. As we show, such *a priori* information can be naturally incorporated in the signal encoding and estimation algorithms by using average rather than worst-case performance metrics to design these systems.

Finally, in Section 3.3 we examine another important extension of the static-case estimation problem of Chapter 2 where, in addition to the signal parameter of interest, the sensor noise power level is unknown. We show how the performance measures and the associated systems we developed in Chapter 2 can be extended to encompass this important case. In the context of all these extensions we will focus our attention on the special case that the sensor noise is IID zero-mean Gaussian, although in most cases our results can be generalized to a much broader class of nonGaussian sensor noises.

3.1 Multiple Sensors

In this section we examine a network generalization of the single-sensor problem stated in (2.1)–(2.2), namely, estimating an unknown parameter A from observation of

$$y_\ell[n] = F_\ell(A + v_\ell[n] + w_\ell[n]) \qquad n = 1, 2, \cdots, N, \quad \ell = 1, 2, \cdots, L \tag{3.1}$$

where $F_\ell(\cdot)$ is a M_ℓ-quantizer of the form (2.2) with thresholds $X_{\ell,1}, \cdots, X_{\ell,M_\ell-1}$, the $v_i[n]$'s are IID processes, and the $w_i[n]$'s denote the applied control input sequences. We use for convenience $\mathbf{Y}^N$ to denote the following $(N\,L) \times 1$ vector of N encoded observations from each of the L sensors

$$\mathbf{Y}^N \triangleq \left[\left[\mathbf{y}_1^N\right]^T \quad \left[\mathbf{y}_2^N\right]^T \quad \cdots \quad \left[\mathbf{y}_L^N\right]^T \right]^T . \tag{3.2}$$

Networks employing encodings in the form of (3.1) provide attractive models for a

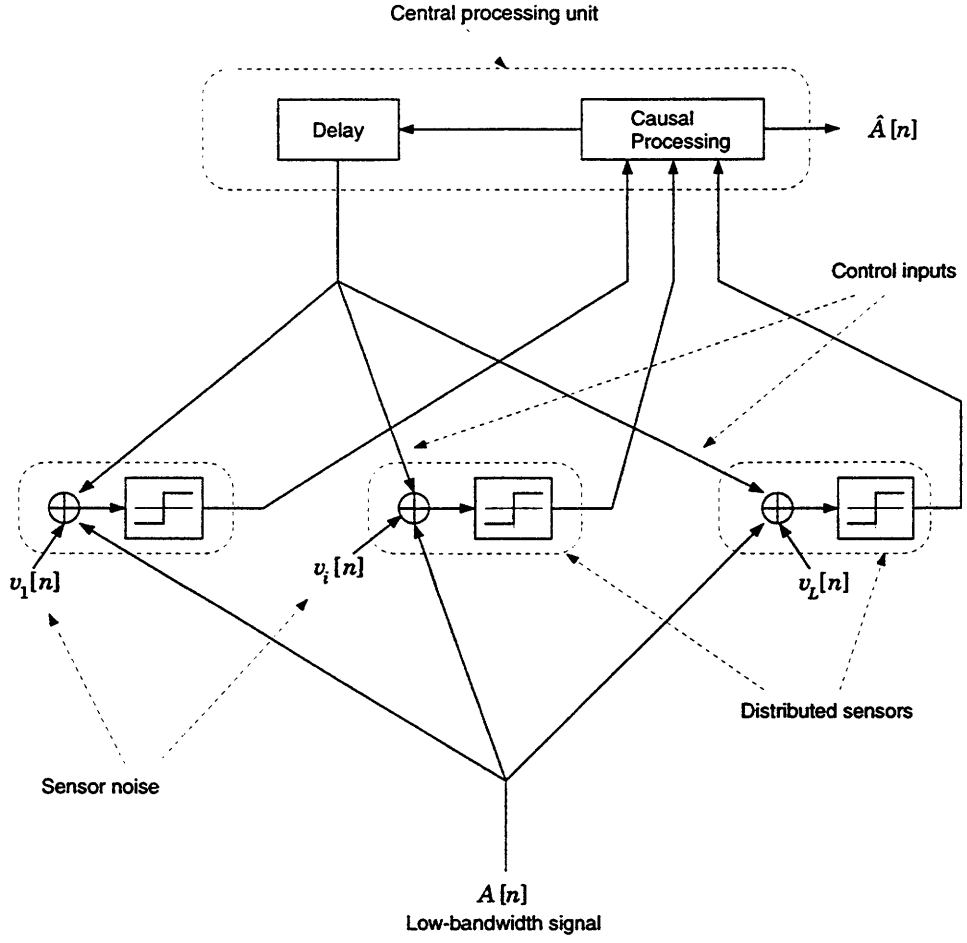


Figure 3-1: Block diagram of a network of distributed signal quantizers using feedback in the context of signal estimation.

number of distributed sensor networks. In Fig. 3-1, for instance, we show the block diagram of a special case of such a distributed estimation network which employs feedback in the selection of the control inputs. In this section, we consider distributed estimation networks with and without feedback.

3.1.1 Statistically Independent Sensor Noises

In the case that the sensor noise processes $v_\ell[n]$ in (3.1) are statistically independent, straightforward extensions of the single-sensor systems developed in Chapters 2 yield network generalizations. In particular, these networks can be analyzed by means of the tools developed for the single-sensor case.

For the remainder of this section we restrict our attention to IID Gaussian sensor noise, which we use as a representative example to illustrate the extensions of the single-sensor

results to the associated multi-sensor settings. Analogous extensions can be similarly derived for all the other scenarios we developed in Sections 2.2–2.3.

Pseudo-noise Control Inputs

We may consider a network of sensors employing encodings of the form (3.1) for which the control inputs are IID pseudo-noise sequences with known statistical description $w_\ell[n] \sim \mathcal{N}(0, \sigma_{w_\ell}^2)$, and which can be adequately modeled as statistically independent of one another and of the sensor noises. We will consider two cases which differ in terms of whether the sensor noise levels and the quantizers are identical or different.

In the case that all quantizers are identical, *i.e.*, $F_\ell(x) = F(x)$ for all x and the sensor noises have equal strength, *i.e.*, $\sigma_{v_\ell} = \sigma_v$ in (3.1), the collection of L observation vectors $\{\mathbf{y}_i^N\}$ can be viewed as a single $(N\,L) \times 1$ observation vector $\mathbf{Y}^N$ collected from a *single* sensor. Hence, in this case all the analysis of Sections 2.2.1 and 2.3.1 applies intact. For instance, in the special case $M_\ell = M = 2$, the optimal noise level is given by $\sigma_{w_\ell} = \sigma_w^{\text{opt}}$ from (2.18) and the associated ML estimator is given by (2.43) where the $N \times 1$ observation vector $\mathbf{y}^N$ is replaced with the $(N\,L) \times 1$ vector $\mathbf{Y}^N$, *i.e.*,

$$\hat{A}_{\text{ML}}\left(\mathbf{Y}^N; \infty\right) = -\sigma_\alpha\, Q^{-1}\left(\frac{\mathcal{K}_1\left(\mathbf{Y}^N\right)}{N\,L}\right) . \tag{3.3}$$

In the general case where the quantizers are distinct and the overall noise levels (summarizing the effects of the sensor noise and the pseudo-noise component) have different strengths, Cramér-Rao bounds and corresponding ML estimators can be formed with minor modifications of the single-sensor problem. Specifically, the optimal pseudo-noise power level σ_{w_ℓ} to be used at the ℓth sensor can be selected so as to optimize the single-sensor performance; for instance for any ℓ for which $M_\ell = 2$, the optimal σ_{w_ℓ} is given by (2.18) with σ_v replaced by σ_{v_ℓ}. Similarly, the ML estimator of A from observation of $\mathbf{Y}^N$ is given by the following extension of the single-sensor EM algorithm

$$\hat{A}_{\text{EM}}^{(k+1)} = \mathcal{I}_\Delta\left(\hat{A}_{\text{EM}}^{(k)} + \sum_{\ell=1}^{L}\sum_{m=1}^{M_\ell} \frac{\sigma_{\alpha_\ell}\,\mathcal{K}_{Y_m}(\mathbf{y}_\ell^N)}{\sqrt{2\,\pi}\,N\,L}\, \frac{\exp\left(-\frac{\left(X_{\ell,m-1}-\hat{A}_{\text{EM}}^{(k)}\right)^2}{2\,\sigma_{\alpha_\ell}^2}\right) - \exp\left(-\frac{\left(X_{\ell,m}-\hat{A}_{\text{EM}}^{(k)}\right)^2}{2\,\sigma_{\alpha_\ell}^2}\right)}{Q\left(\frac{X_{\ell,m-1}-\hat{A}_{\text{EM}}^{(k)}}{\sigma_{\alpha_\ell}}\right) - Q\left(\frac{X_{\ell,m}-\hat{A}_{\text{EM}}^{(k)}}{\sigma_{\alpha_\ell}}\right)}\right), \tag{3.4}$$

where

$$\sigma_{\alpha_\ell} = \sqrt{\sigma_{v_\ell}^2 + \sigma_{w_\ell}^2} \, .$$

As a direct extension of the single-sensor results discussed in Chapter 2, at high peak SNR (*i.e.*, $\Delta \gg \sigma_{v_\ell}$ for $\ell = 1, 2, \cdots, L$) by selecting the pseudo-noise levels as $\sigma_{w_\ell} \sim \lambda \Delta$, the worst-case information loss can be made to grow as slow as the square of the parameter dynamic range Δ, for any network of fixed size L with a fixed set of quantizers, and sensor noise components with fixed statistical characterization.

Known Control Inputs

Similarly, in the case that the control inputs $w_\ell[n]$ in (3.1) are known for estimation we can easily extend the associated encoding strategies and estimation algorithms so as to achieve optimal performance in terms of minimizing the worst-case information loss rate. Specifically, we may design the encoding strategy used at each sensor separately viewing it as a single-sensor system. If, for instance, $M_\ell = M = 2$, we can select the control input sequence at the ℓth sensor according to (2.26) where K_ℓ is given by (2.27) where χ is replaced by

$$\chi_\ell = \frac{\Delta}{\sigma_{v_\ell}} \, , \tag{3.5}$$

$\lambda = 2/d_\infty$, and d_∞ is given by (2.30).

The performance of networks of sensors in the context of known control inputs is a natural extension of the associated single sensor performance; for a fixed network size, with fixed quantizers and sensor noise PDFs, the worst-case information loss can be made to grow linearly with the signal dynamic range Δ by means of the encoding scheme described by (3.5).

Natural extensions of the EM algorithm (2.49) can be used to perform efficient data fusion and signal estimation in the multi-sensor case. In particular, assuming that a K_ℓ-periodic sequence $w_\ell[n]$ is used as the bias of the ℓth quantizer $F_\ell(\cdot)$ in (3.1), we have

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)} + \sum_{\ell=1}^{L} \sum_{\substack{1\le i\le K_\ell \\ 1\le m\le M_\ell}} \frac{\sigma_{v_\ell} \mathcal{K}_{Y_m}(\mathbf{y}_\ell^{\bar{N}_\ell}[i])}{\sqrt{2\pi N}} \frac{\exp\left(-\frac{\left(z_\ell^{(k)}[m-1,i]\right)^2}{2}\right) - \exp\left(-\frac{\left(z_\ell^{(k)}[m,i]\right)^2}{2}\right)}{Q\left(z_\ell^{(k)}[m-1,i]\right) - Q\left(z_\ell^{(k)}[m,i]\right)}\right), \tag{3.6a}$$

and

$$\hat{A}_{\mathrm{ML}} = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}, \tag{3.6b}$$

where $\bar{N}_\ell = \lfloor N/K_\ell \rfloor$, $\mathbf{y}^{\bar{N}_\ell}[i]$ is the $\bar{N}_\ell \times 1$ vector comprised of the elements of the ℓth K_ℓ-decimated subsequence, *i.e.*,

$$\mathbf{y}_\ell^{\bar{N}_\ell}[i] \triangleq \left[y_\ell[i] \quad y_\ell[K_\ell + i] \quad \cdots \quad y_\ell[N - K_\ell + i]\right]^T \qquad i = 1, 2, \cdots, K_\ell, \tag{3.6c}$$

and

$$z_\ell^{(k)}[m,i] = \frac{X_{\ell,m} - \hat{A}_{\mathrm{EM}}^{(k)} - w_\ell[i]}{\sigma_{v_\ell}}. \tag{3.6d}$$

If all the quantizers are identical, *i.e.*, $F_\ell(\cdot) = F(\cdot)$, and all the sensor noises have identical PDFs, the above encoding design method results in selecting the same control input sequence for each sensor, *i.e.*, $w_\ell[n] = w[n]$ from (2.26) where K is given by (2.27). When in addition $L \gg K$, or when L is an integer multiple of K, the encoding strategy can be simplified even further by *spatially* distributing the K possible control input values. Specifically, consider for simplicity the case where $n = L/K$ is an integer. By dividing the L sensors into K groups of n sensors, and by setting the control input of any sensor within a given group equal to one of the K distinct samples of the K-periodic sequence (2.26) for all n, we can achieve optimal encoding performance without the need for a time-varying control input.

Control Inputs in the Presence of Feedback

Networks exploiting feedback from the host (observing the quantized outputs) to the sensors (using quantizer bias control) can also be analyzed using the associated single-sensor

principles. As a natural extension of the single sensor results, at each sensor our objective is to operate around the point where the information loss is minimized; in the case $M = 2$ for instance, performance is optimized when we operate in the vicinity of the quantizer threshold. As a generalization of the single-sensor analysis for quantizer bias control via feedback, the control input can be selected using (2.52), where $\hat{A}[n-1]$ denotes the estimate of A based on observations collected from all L sensors up to and including time $n-1$. For instance, in the case $M = 2$, the multi-sensor extension of the ML estimator (2.53) is given by

$$\hat{A}_{\mathrm{EM}}^{(k+1)}[n] = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)}[n] + \sum_{m=1}^{n}\sum_{\ell=1}^{L} y_{\ell}[m]\,\frac{\sigma_{v_\ell}}{\sqrt{2\,\pi}\,n\,L}\,\frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{ML}}[m-1]-\hat{A}_{\mathrm{EM}}^{(k)}[n]\right)^2}{2\,\sigma_{v_\ell}^2}\right)}{Q\left(y_{\ell}[m]\,\frac{\hat{A}_{\mathrm{ML}}[m-1]-\hat{A}_{\mathrm{EM}}^{(k)}[n]}{\sigma_{v_\ell}}\right)}\right), \tag{3.7a}$$

initialized with $\hat{A}_{\mathrm{EM}}^{(0)}[n] = \hat{A}_{\mathrm{ML}}[n-1]$ and $\hat{A}_{\mathrm{ML}}[0] = 0$, where for any n,

$$\hat{A}_{\mathrm{ML}}[n] = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}[n]\,. \tag{3.7b}$$

The associated multi-sensor extension of (2.58) is similarly given by

$$\check{A}[n] = \begin{cases} \hat{A}_{\mathrm{ML}}[n] \quad \text{from (3.7)} & \text{if } n \le n_o \\ \mathcal{I}_{\Delta}\left(\check{A}[n-1] + \sqrt{\frac{\pi}{2}}\,\frac{\sum_{\ell=1}^{L}\sigma_{v_\ell}^{-1}\,y_{\ell}[n]}{n\,\sum_{\ell=1}^{L}\sigma_{v_\ell}^{-2}}\right) & \text{if } n > n_o \end{cases}, \tag{3.8}$$

Fig. 3-2 depicts the MSE performance of the ML estimator (3.7) and $\check{A}[N]$ given by (3.8) for a network of $L = 5$ sensors. As in the single-sensor case, the MSEs of both estimators practically achieve the associated Cramér-Rao bound corresponding to a 2 dB information loss for moderate N. In general, spatial redundancy (large L) leads to faster convergence to the associated 2 dB bound. This fact is exploited in Chapter 5 where we develop encodings for sensor networks used to estimate fast time-varying signals.

In the Gaussian scenario, for networks of sensors encoding 1 bit of information per measurement and employing quantizer bias control with feedback, the associated information loss can be directly obtained using appropriate interpretation of Fig. 2-7 describing the single-sensor case. Similar extensions of the associated single-sensor problem can be

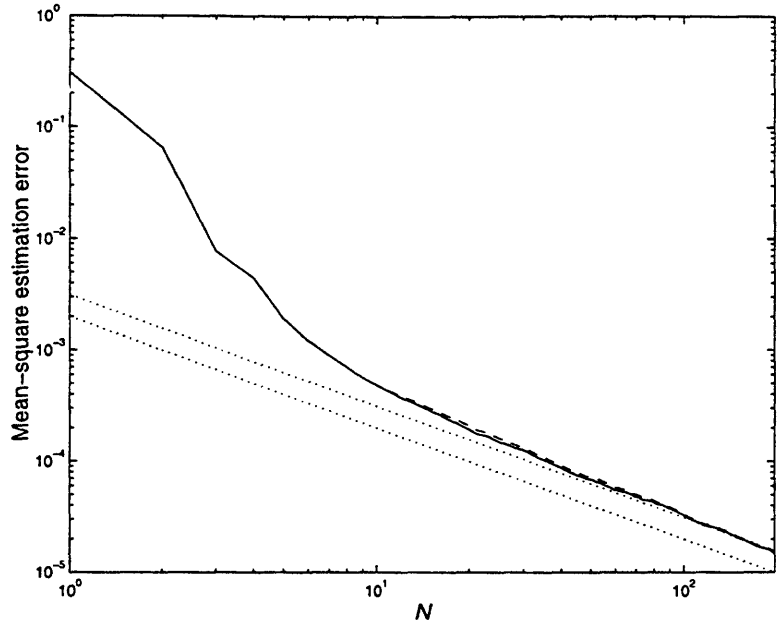


Figure 3-2: MSE for $\hat{A}_{\mathrm{ML}}[N]$ and $\check{A}[N]$ for a network of $L = 5$ two-level quantizers, using feedback in the selection of the control input, and associated Cramér-Rao bounds (see also caption of Fig. 2-9). The sensor noise levels are 0.08, 0.08 0.08, 0.2, and 0.4, while $A = 0.4$ and $\Delta = 1$.

obtained for any set of sensor noises for $M \geq 2$. For instance, if feedback is available and properly used in the multi-sensor setting shown in Fig. 3-1, a small worst-case information (and MSE) loss can be achieved, independent of the dynamic range and the noise power levels. This small information loss, however, will in general depend on both the quantizers $F_\ell(\cdot)$ and the sensor noise PDFs.

3.1.2 Perfectly Correlated Sensor Noises

In this section we consider an example involving sensor noises that are spatially correlated. In particular, we consider the case where the concurrent sensor noise samples are spatially perfectly correlated, *i.e.*, $v_\ell[n] = v[n]$ for $1 \leq \ell \leq L$ and where $v[n]$ is an IID sequence.

This model may be naturally suited for distributed estimation settings in which there is an additive distortion component that is identical at a number of distinct sensors. In addition, this model may arise in a variety of other applications involving estimation from coarsely digitized measurements; in the context of analog to digital conversion of noisy signals for instance, this model may provide a reasonably accurate representation of the noisy analog signal that is to be digitized by each element in an A/D converter array of

inexpensive components.

For such systems, the analysis in the presence of known periodic control inputs, or of control inputs selected using feedback information, naturally decouples to that of the associated single-sensor problems we have already considered in Chapter 2. For instance, a network of L binary quantizers where the control inputs used are known for estimation, is equivalent to a single $L+1$-level sensor with known time-varying thresholds.

Henceforth, we focus on the special case where the control inputs $w_\ell[n]$ correspond to pseudo-noise sequences that are well modeled as independent IID Gaussian sequences, each with variance σ_w^2, and focus on the case $M = 2$.

Motivated by the form of the estimator (3.3) we focus on the following elementary estimators of A

$$\hat{A}\left(\mathbf{Y}^N\right) = \mathcal{I}_\Delta\left(-\sigma_\alpha Q^{-1}\left(\tilde{k}_1\left(\mathbf{Y}^N\right)\right)\right) \tag{3.9a}$$

where $\mathbf{Y}^N$ is given by (3.2) and

$$\tilde{k}_1\left(\mathbf{Y}^N\right) = \frac{\mathcal{K}_1\left(\mathbf{Y}^N\right)}{NL} = \frac{1}{2} + \frac{1}{NL}\sum_{n=1}^{N}\sum_{\ell=1}^{L} y_\ell[n]\,. \tag{3.9b}$$

We will mainly focus on the case where $\sigma_v \ll \Delta$, which corresponds to significant worst-case information loss in the case $L = 1$ examined in Chapter 2, even when the pseudo-noise level is optimally selected (see dash-dot curve in Fig. 2-7 for large Δ/σ_v) . We can show by methods very similar to those used in App. A.4, that for large N and L, the MSE of the estimator in (3.9) is reasonably approximated as follows

$$E\left[\left(A - \hat{A}\left(\mathbf{Y}^N\right)\right)^2\right] \approx \frac{1}{NL}\mathcal{B}_{\mathcal{N}}\left(A;\ \sigma_\alpha\right) + \frac{L-1}{NL}\,\eta(A, \sigma_v, \sigma_w) \tag{3.10}$$

where $\mathcal{B}_{\mathcal{N}}\left(A;\ \sigma_\alpha\right)$ is given by (2.16), and

$$\eta(A, \sigma_v, \sigma_w) = \rho(A, \sigma_v, \sigma_w) - Q\left(-\frac{A}{\sigma_\alpha}\right)^2\,,$$

where

$$\rho(A, \sigma_v, \sigma_w) = \frac{1}{2\pi\,\sigma_v}\int_{v=-\infty}^{\infty} Q\left(-\frac{A+v}{\sigma_w}\right)^2 \exp\left(-\frac{v^2}{2\,\sigma_v^2}\right)\,dv\,.$$

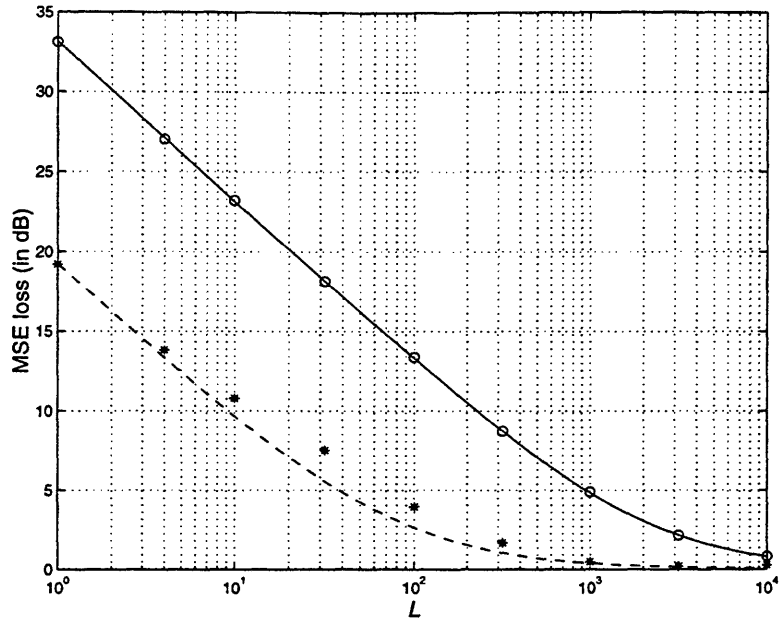


Figure 3-3: Estimation in the presence of perfectly correlated sensor noise components. The pseudo-noise sequences $w_i[n]$ for $i = 1, 2, \cdots, L$ are modeled as independent IID Gaussian noise sources, independent of $v[n]$, with $\sigma_w = 0.6$. The solid (dashed) curve corresponds to the predicted MSE loss, while the "o" ("*") marks depict the MSE Loss from Monte-Carlo simulations for the estimator (3.9) for $A = 0.6$ and $\sigma_v = 0.02$ ($\sigma_v = 0.1$).

In Fig. 3-3 we present the results of Monte-Carlo simulations on the MSE loss of the estimator (3.9) for two representative sensor noise levels. The solid (dashed) curve depicts the MSE estimate (3.10) for $A = 0.6$, $\sigma_v = 0.02$ ($\sigma_v = 0.1$), while the "o" ("$*$") symbols depict the associated simulated MSE from Monte-Carlo simulations. The pseudo-noise level used at all sensors was $\sigma_w = 0.6$, while the signal dynamic range is $\Delta = 1$. As the figure illustrates, (3.10) predicts the MSE loss fairly accurately for large L in these two examples.

Eqn. (3.10) suggests a method for obtaining an approximate value for the minimum number of sensors $L_{\min}$ that is required in order to achieve performance within C dB of the unconstrained performance σ_v^2/N, for $\sigma_v^2 \ll \Delta$, by means of the estimator (3.9). In this parameter regime, $\mathcal{B}_{\mathcal{N}}(A;\ \sigma_\alpha) \gg \eta(A, \sigma_v, \sigma_w)$ and $\sigma_w \approx \sigma_\alpha$, which together imply that

$$L_{\min}\,\sigma_v^2 \approx \frac{\mathcal{B}_{\mathcal{N}}(\Delta;\ \sigma_w)}{10^{C/10} - 1}\ . \tag{3.11}$$

In Fig. 3-4 we present the network size $L_{\min}$ required to achieve MSE performance within

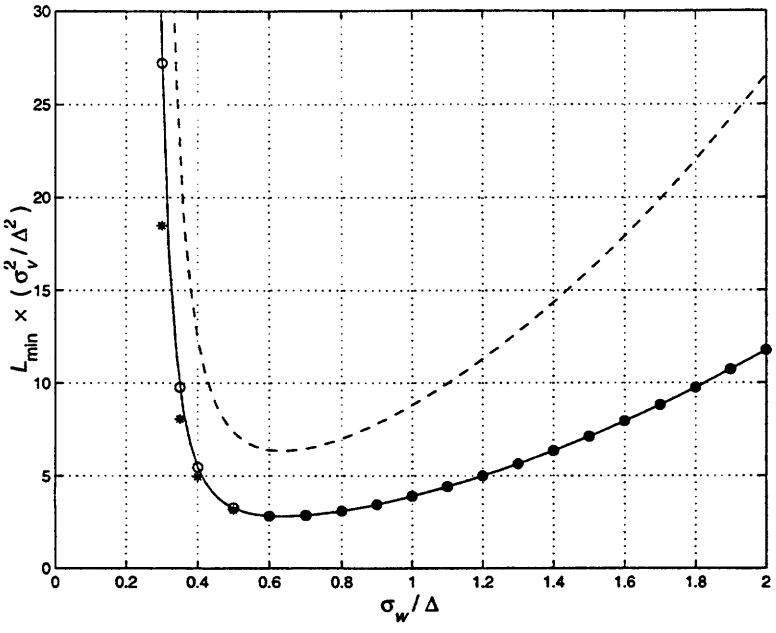


Figure 3-4: Minimum network size $L_{\min}$ required for reaching within 2 dB (solid curve) and 1 dB (dashed curve) of the infinite-resolution MSE, as predicted by (3.11). The "o" and "$*$" marks depict the required $L_{\min}$ according to (3.10) for $\sigma_v/\Delta = 0.02$ and $\sigma_v/\Delta = 0.1$, respectively.

2 dB (solid curve) and 1 dB (dashed curve) of the best performance based on the original infinite-resolution noisy measurements (*i.e.*, $\mathcal{B}\left(A;\, \mathrm{S}^N\right) = \sigma_v^2/N$), as a function of σ_w/Δ according to (3.11). The "o" and "$*$" marks depict the required network size $L_{\min}$ by means of (3.10) for $\sigma_v/\Delta = 0.02$ and $\sigma_v/\Delta = 0.1$, respectively. Note for instance that, if $\Delta = 1$, then for $\sigma_w = 0.6$ and $\sigma_v = 0.1$, a network of size $L \approx 280$ is needed to achieve the infinite-resolution bound within 2 dB, while a network of size $L \approx 635$ reaches within 1 dB of the infinite-resolution performance.

3.2 Incorporation of Prior Information

In a number of applications involving estimation from digital encodings of the form of quantizer bias control, we may have prior information about the relative likelihood of various values of the information-bearing signal, which in the static case is a single parameter A. Such *a priori* information can arise from a variety of sources, such as the underlying mechanisms that generate the information-bearing signal. As we will see in Chapter 5, temporal correlations in the information bearing signal can often be exploited in the form

of *a priori* information.

In all these cases an average rather that worst-case performance metric is more naturally suited for system analysis and design. As we show in this section, we can design encodings based on quantizer bias control for which the average information loss rates exhibit strikingly similar behavior to the associated worst-case information loss rates developed for unknown parameter estimation. Specifically, the quality of the encoding is characterized by the average encoding performance given by

$$\bar{\mathcal{B}}(A;\ y) \triangleq E[\mathcal{B}(A;\ y)] = \int_A \mathcal{B}(A;\ y)\ p_A(A)\ dA\ . \tag{3.12}$$

and the average information loss given by

$$\bar{\mathcal{L}}(p_A(\cdot)) \triangleq E[\mathcal{L}(A)] = \int_A \mathcal{L}(A)\, p_A(A)\ dA\ , \tag{3.13}$$

where $\mathcal{L}(A)$ is given by (2.4), and $\mathcal{B}(A;\ y)$ is the Cramér-Rao bound for estimating an unknown parameter A based on any one sample of the IID sequence $y[n]$. Since the best possible MSE performance based on the uncoded set of observations $s[n]$ satisfies

$$\bar{\mathcal{B}}(A;\ s) = \mathcal{B}(A;\ s)\ ,$$

the average information loss (3.13) and average Cramér-Rao bound (3.12) can be used interchangeably as measures of performance.

The metrics (3.12) and (3.13) are reasonable performance metrics for assessing the encoding performance for a large number of observations, and, in particular, as $N \to \infty$. Specifically, if N is large enough so that the information due to the encodings $\mathbf{y}^N$ dominates the information from the prior information, (3.12) and (3.13) represent the MSE limits for estimation based on the encodings only, averaged with respect to the prior $p_A(\cdot)$. At the other extreme where $N = 0$, there is no information loss in terms of using the encodings instead of the original data: in both cases the only information available is due to the prior information. In general, the larger N the more the information due to the encodings, and thus the larger the information loss. Thus, for small finite N the information loss due to the encodings is in general less than (3.13).

We next consider two representative forms of *a priori* information. First, we consider

the case where the random variable A is uniformly distributed within the range $(-\Delta, \Delta)$. Consequently, we consider a case where the parameter A is a Gaussian random variable (and thus is not range-limited).

3.2.1 Uniformly Distributed Signal

In this section we develop extensions of the encoding and estimation algorithms of Chapter 2 where the objective is to optimize average rather than worst-case performance over the signal dynamic range. It is often reasonable to assume that the random variable A is *a priori* uniformly distributed in $(-\Delta, \Delta)$.

Estimation Algorithms

Analogously to the ML estimate for unknown parameters, we may consider the maximum a posteriori (MAP) estimate of the random variable A given $\mathbf{y}^N$ [31], namely

$$\hat{A}_{\text{MAP}}\left(\mathbf{y}^N\right) = \arg\max_{\theta} \left[\ln\left(p_{\mathbf{y}^N|A}\left(\mathbf{y}^N|\theta\right)\right) + \ln(p_A(\theta))\right] .$$

Due to the particular form of the prior $p_A(\cdot)$, for any type of encodings generated via quantizer bias control, the MAP estimate is *identical* the associated ML estimate with range restriction in (Δ, Δ) developed in Chapter 2. Consequently, for all encoding scenarios of the form of quantizer bias control that we have considered in Chapter 2, the associated estimation algorithms we have developed in Chapter 2 are asymptotically optimal, in the sense that they asymptotically achieve the associated average information loss (3.12) of the encodings. Consequently, we simply need to redesign the encoding strategies having in mind that we now need to optimize over average rather than worst-case information loss performance.

Pseudo-noise inputs

We first consider the estimation problem (2.1) with $F(\cdot)$ given by (2.2), where the control input is a pseudo-noise sequence. We assume that $w[n]$ and $v[n]$ are independent IID zero-mean Gaussian sequences with variance σ_v^2, and σ_w^2 respectively, and independent of A. We wish to select the pseudo-noise level σ_w so as to minimize the average information loss (3.13). For convenience, we use $\bar{\mathcal{B}}(\Delta, \sigma_v, \sigma_w)$ to denote (3.12) for a given Δ, σ_v, and

σ_w, and $\bar{\mathcal{B}}(\Delta,\ \sigma_\alpha)$ to denote (3.12) for a given Δ and $\sigma_\alpha = \sqrt{\sigma_v^2 + \sigma_w^2}$. Similarly, we let $\bar{\mathcal{L}}(\Delta,\ \sigma_v,\ \sigma_w)$ denote (3.13) for a given Δ, σ_v, and σ_w, and $\bar{\mathcal{L}}(\Delta,\ \sigma_\alpha)$ denote (3.13) for a given Δ and $\sigma_\alpha = \sqrt{\sigma_v^2 + \sigma_w^2}$.

For any admissible sensor noise distribution, let $\sigma^{\text{opt}}(\Delta)$ denote the sensor noise level that minimizes the average encoding performance (3.12), *i.e.*,

$$\sigma^{\text{opt}}(\Delta) = \arg\min_{\sigma}\ \bar{\mathcal{B}}(\Delta,\ \sigma)\ . \tag{3.14}$$

In a manner analogous to the treatment of the unknown parameter case we can show that $\sigma^{\text{opt}}(\Delta) > 0$ for any $\Delta > 0$. In particular,

$$\sigma^{\text{opt}}(\Delta) = \arg\min_{\sigma}\ \bar{\mathcal{B}}(\Delta,\ \sigma) = \Delta\ \arg\min_{\sigma}\ E\left[\mathcal{B}(A/\Delta;\ \sigma_v/\Delta)\right] = \sigma^{\text{opt}}(1)\,\Delta\ , \tag{3.15}$$

which also implies that

$$\bar{\mathcal{B}}\left(\Delta,\ \sigma^{\text{opt}}(\Delta)\right) = \Delta^2\,\bar{\mathcal{B}}\left(1,\ \sigma^{\text{opt}}(1)\right)\ . \tag{3.16}$$

Fig. 3-5 depicts $\bar{\mathcal{B}}(\Delta,\ \sigma_v)\,/\Delta^2$ as a function of σ_v/Δ for A being uniformly distributed. As the figure reveals, $\sigma^{\text{opt}}(\Delta) > 0$ for $\Delta > 0$. In particular, numerical evaluation of (3.14) for $\Delta = 1$ together with (3.15) yields

$$\sigma^{\text{opt}}(\Delta) = \sigma^{\text{opt}}(1)\,\Delta \approx 0.4622\,\Delta\ . \tag{3.17}$$

The existence of a nonzero optimal noise level in terms of minimizing the average MSE performance of the encoding for any given Δ can be exploited to provide encoding performance benefits by means of pseudo-noise bias control. Specifically, by choosing the pseudo-noise power level as

$$\sigma_w^{\text{opt}}(\Delta) = \begin{cases} \sqrt{[\sigma^{\text{opt}}(\Delta)]^2 - \sigma_v^2} & \text{if } \sigma^{\text{opt}}(\Delta) > \sigma_v \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

where $\sigma^{\text{opt}}(\Delta)$ is given by (3.17), the average encoding performance is given by (3.16) at

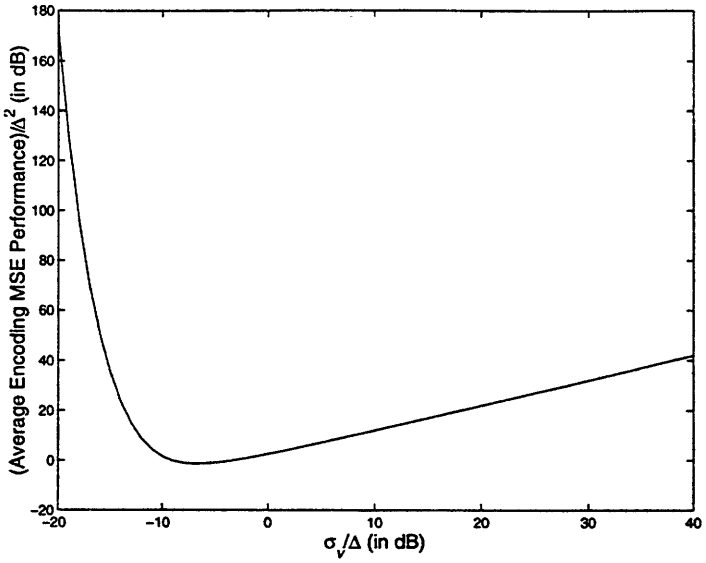


Figure 3-5: $\bar{B}(\Delta,\ \sigma)/\Delta^2$ as a function of σ_v/Δ when A is *a priori* uniformly distributed in $(-\Delta,\ \Delta)$.

high SNR χ defined as $\chi = \Delta/\sigma_v$, which, in conjunction with (2.8), gives

$$\bar{\mathcal{L}}^{\text{pn}}(\chi) = \bar{\mathcal{L}}\left(1,\ \sigma^{\text{opt}}(1)\right)\ \chi^2 \tag{3.19}$$

for large enough χ. For comparison, consider the case where the control input is set to zero and where the sensor noise is Gaussian. The average information loss in this case is given by

$$\begin{aligned}\bar{B}_{\mathcal{N}}^{\text{free}}(\Delta;\ \sigma_v) &= \frac{1}{2\Delta}\int_{A=-\Delta}^{\Delta} Q\left(\frac{\Delta}{\sigma_v}\right) Q\left(-\frac{\Delta}{\sigma_v}\right) \frac{1}{\sqrt{2\,\pi}\,\sigma_v}\exp\left(\frac{\Delta^2}{\sigma_v^2}\right) dA \\ &\approx \frac{\sigma_v^2}{\sqrt{2\,\pi}\chi^3}\exp\left(\chi^2/2\right)\end{aligned}$$

where the approximation holds for large χ. Combining the above approximation with (2.8) we obtain the average information loss in Gaussian noise in the absence of a control input, namely,

$$\bar{\mathcal{L}}^{\text{free}}(\chi) \approx \frac{1}{\sqrt{2\,\pi}\chi^3}\exp\left(\chi^2/2\right)\ ,$$

which grows at a rate much faster that the optimal pseudo-noise case in (3.19). Again, we can easily show by extending the proof of the unknown parameter case that pseudo-noise

yields performance benefits for any admissible sensor noise PDF and any quantizer with $M \geq 2$.

Known Control Inputs and Control Inputs via Feedback

Similar extensions of the unknown noise counterparts can be developed for known control inputs; we may use periodic control inputs of the form (2.26), where by choosing K according to (2.26) we can again achieve a linear growth rate for the average information loss as a function of χ.

Finally, in the presence of feedback the encoding and estimation strategies used in Chapter 2 achieve the 2 dB loss for all parameter values, which implies that worst-case and average performance are in this case identical.

Similar behavior is exhibited by other *a priori* PDFs. We next describe encoding strategies and estimation algorithms in the case that the random variable A is Gaussian and where the sensor noise is also Gaussian.

3.2.2 Normally Distributed Signal

In a number of practical scenarios the information-bearing signal is not range-limited, that is, a uniformly distributed PDF description fails to provide an accurate description of the *a priori* parameter characterization. Often, it is reasonable to assume that A is *a priori* normally distributed with mean m_A and power level σ_A^2. For instance, this is a naturally suited *a priori* signal description in cases where the random parameter denotes the overall effect of large collections of finite power events. As we will show, there is also a natural measure of signal-to-noise ratio in the design and performance evaluation of these systems. Again we focus on the case that the sensor noise PDF is Gaussian.

Pseudo-Noise Control Inputs

We first consider the estimation problem (2.1) with $F(\cdot)$ given by (2.2) in the case that the control input is a pseudo-noise sequence. We assume that $w[n]$ and $v[n]$ are independent IID zero-mean and normally distributed sequences with variance σ_v^2, and σ_w^2 respectively, and independent of the Gaussian random variable A. We wish to select the pseudo-noise power level σ_w so as to minimize the average information loss (3.13). For illustration, we focus on the case $M = 2$.

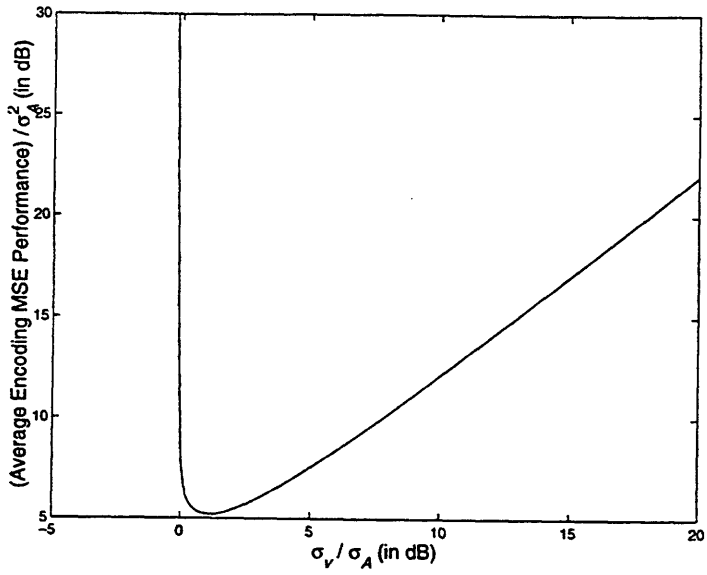


Figure 3-6: $\bar{\mathcal{B}}(\sigma_A,\ \sigma)/\sigma_A^2$ as a function of σ_v/σ_A when A is *a priori* zero-mean and normally distributed with variance σ_A^2.

First we consider the special case where A is zero-mean. For convenience, we use $\bar{\mathcal{B}}(\sigma_A,\ \sigma_v,\ \sigma_w)$ to denote (3.12) for a given σ_A, σ_v and σ_w, and $\bar{\mathcal{B}}(\sigma_A,\ \sigma_\alpha)$ denote (3.12) for a given σ_A and $\sigma_\alpha = \sqrt{\sigma_v^2+\sigma_w^2}$. Similarly, let $\bar{\mathcal{L}}(\sigma_A,\ \sigma_v,\ \sigma_w)$ denote (3.13) for a given σ_A, σ_v, and σ_w, and $\bar{\mathcal{L}}(\sigma_A,\ \sigma_\alpha)$ denote (3.13) for a given σ_A and $\sigma_\alpha = \sqrt{\sigma_v^2+\sigma_w^2}$. Following the analysis of Section 3.2.1 we can show that for any given signal power level $\sigma_A > 0$, there is an optimal aggregate noise level σ^{opt} in terms of minimizing the average encoding loss (3.12). In particular, similar to (3.15) we have

$$\sigma^{\text{opt}}(\sigma_A) = \sigma^{\text{opt}}(\sigma_1)\,\sigma_A \approx 1.1395\,\sigma_A \tag{3.20}$$

where $\sigma^{\text{opt}}(\sigma_1)$ has been numerically computed. Fig. 3-6 depicts $\bar{\mathcal{B}}(\sigma_A,\ \sigma_v)/\sigma_A^2$ as a function of σ_v/σ_A, where $A \sim \mathcal{N}(0, \sigma_A^2)$. As the figure reveals, $\sigma^{\text{opt}}(\sigma_A) > 0$ for $\sigma_A > 0$. Eqn. (3.20) also implies that

$$\bar{\mathcal{B}}\left(\sigma_A,\ \sigma^{\text{opt}}(\sigma_A)\right) = \sigma_A^2\,\bar{\mathcal{B}}\left(1,\ \sigma^{\text{opt}}(1)\right). \tag{3.21}$$

We can exploit (3.21) to show that the average information loss (3.13) can be made to grow as slow as quadratically with an appropriately defined measure of SNR; by choosing the

pseudo-noise power level as

$$\sigma_w^{\text{opt}}(\sigma_A) = \begin{cases} \sqrt{[\sigma^{\text{opt}}(\sigma_A)]^2 - \sigma_v^2} & \text{if } \sigma^{\text{opt}}(\sigma_A) > \sigma_v \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

where $\sigma^{\text{opt}}(\sigma_A)$ is given by (3.20), the average encoding performance is given by (3.21) which, in conjunction with (2.8), and by letting $\bar{\chi} = \sigma_A/\sigma_v$ gives

$$\bar{\mathcal{L}}^{\text{pn}}(\bar{\chi}) = \bar{\mathcal{L}}\left(1, \sigma^{\text{opt}}(1)\right) \bar{\chi}^2 \tag{3.23}$$

for $\bar{\chi} \geq 1/\sigma^{\text{opt}}(1)$. For comparison, performance degrades rapidly at high SNR $\bar{\chi}$ if $w[n] = 0$;

$$\bar{\mathcal{L}}^{\text{free}}(\bar{\chi}) = \frac{1}{\bar{\chi}} \int \frac{Q(u)\, Q(-u)}{f^2(u)} f\left(\frac{u}{\bar{\chi}}\right) du\,, \tag{3.24}$$

which is finite[1] only for $\bar{\chi} < 1$. Furthermore, due to the optimal selection of the pseudo-noise power level in (3.22), we have

$$\bar{\mathcal{L}}^{\text{pn}}(\bar{\chi}) \leq \bar{\mathcal{L}}^{\text{free}}(\bar{\chi}) \tag{3.25}$$

for all $\bar{\chi}$.

For $\mathrm{m}_A \neq 0$, the optimal pseudo-noise level depends both on σ_A and on m_A. For $\sigma_A \ll \mathrm{m}_A$, the random variable A is effectively distributed within a very small region around m_A. In that case the worst-case performance analysis of Chapter 2 can be used to accurately predict average performance (where Δ is replaced with m_A):

$$\begin{aligned} \sigma^{\text{opt}}(\sigma_A, \mathrm{m}_A) &= \arg\min_{\sigma_v} \bar{\mathcal{B}}(\sigma_A, \mathrm{m}_A, \sigma_v) \\ &\approx \arg\min_{\sigma_v} \mathcal{B}(\mathrm{m}_A, \sigma_v) \approx \frac{2}{\pi} |\mathrm{m}_A|, \end{aligned}$$

where $\mathcal{B}(\mathrm{m}_A, \sigma_v)$ is given by the right hand side of (2.16) for $A = \mathrm{m}_A$ and $\sigma_\alpha = \sigma_v$. For

[1]The information loss in the encodings is finite for any $N < \infty$. The fact that $\bar{\mathcal{L}}^{\text{free}}(\sigma_A/\sigma_v)$ diverges for $\sigma_A > \sigma_v$ simply implies that the information loss in the encodings is an increasing function of N with no upper bound.

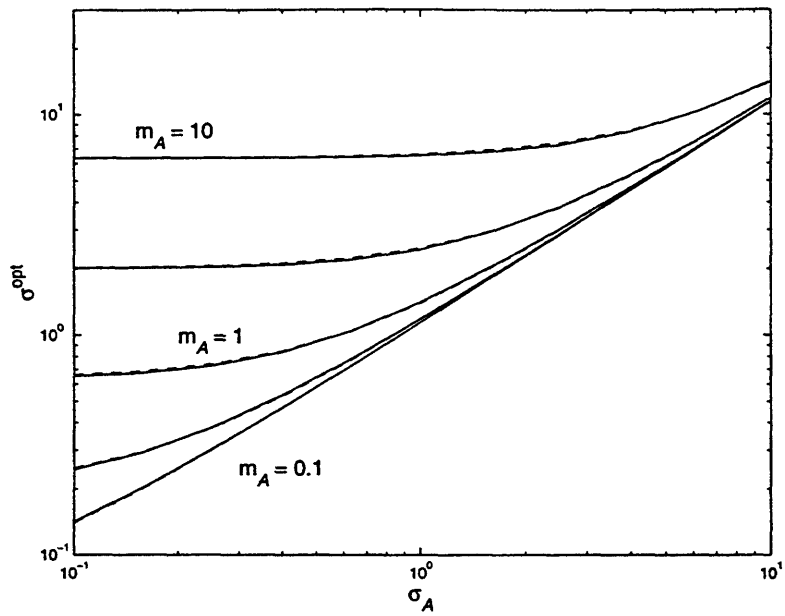


Figure 3-7: Each solid curve depicts the numerically computed value of $\sigma^{\rm opt}(\sigma_A, {\rm m}_A)$ as a function of σ_A for a given ${\rm m}_A$. The dashed curves correspond to the associated predicted values based on (3.26).

arbitrary ${\rm m}_A$ and σ_A, the optimal pseudo-noise level is accurately approximated by

$$\sigma^{\rm opt}(\sigma_A, {\rm m}_A) \approx [(1.1395)^r \, \sigma_A^r + (2/\pi)^r \, |{\rm m}_A|^r]^{\frac{1}{r}} \, , \tag{3.26}$$

with $r = 1.6$. Fig. 3-7 shows the accuracy of the approximation (3.26) (dashed) in terms of predicting the optimal pseudo-noise level $\sigma^{\rm opt}(\sigma_A, {\rm m}_A)$, obtained via numerically optimization (solid).

The (MAP) estimate of A given $\mathbf{y}^N$ can be readily implemented by means of the EM algorithm described in App. A.5; use Eqns. (A.36), (A.37) in conjunction with (A.28) and (A.30). In this case these equations specialize to the following algorithm

$$\hat{A}_{\rm EM}^{(k+1)} = {\rm m}_A + \frac{1}{1+\frac{\sigma_\alpha^2}{N\sigma_A^2}} \left[\hat{A}_{\rm EM}^{(k)} - {\rm m}_A + \right.$$

$$\left. \sum_{m=1}^{M} \frac{\sigma_\alpha \, \mathcal{K}_{Y_m}(\mathbf{y}^N)}{\sqrt{2\pi}\, N} \, \frac{\exp\left(-\frac{\left(X_{m-1}-\hat{A}_{\rm EM}^{(k)}\right)^2}{2\sigma_\alpha^2}\right) - \exp\left(-\frac{\left(X_m-\hat{A}_{\rm EM}^{(k)}\right)^2}{2\sigma_\alpha^2}\right)}{Q\left(\frac{X_{m-1}-\hat{A}_{\rm EM}^{(k)}}{\sigma_\alpha}\right) - Q\left(\frac{X_m-\hat{A}_{\rm EM}^{(k)}}{\sigma_\alpha}\right)} \right], \tag{3.27a}$$

and where $\hat{A}_{\text{MAP}}\left(\mathbf{y}^N\right)$ is given by

$$\hat{A}_{\text{MAP}}\left(\mathbf{y}^N\right) = \lim_{k\to\infty} \hat{A}_{\text{EM}}^{(k)} \,. \tag{3.27b}$$

In general, for large N the MAP estimate (3.27) is approximately given by the ML estimate. Thus asymptotically it achieves the average information loss (3.13).

Known Control Inputs

A priori information can be also incorporated in the system (2.1)–(2.2) in case the control sequence is known for estimation so as to enhance performance. Specifically, optimized encodings together with the corresponding MAP estimators that approach the associated bounds (3.12) can be constructed.

For brevity we again focus on the system corresponding to $M = 2$. We simply need to consider the case where the Gaussian random variable A is zero-mean; given that the control input is known to the estimator, the encoding strategies we develop in this section for the zero-mean case can be readily modified to accommodate the general case.

Unlike the range-limited information-bearing signals considered in Chapter 2 and Section 3.2.1, periodic control inputs are inadequate for achieving optimal performance[2]. For this reason we consider aperiodic control inputs, and in particular control inputs that are sample paths of an IID zero-mean Gaussian random process with power level σ_w^2. The objective is to determine the power level of the control input signal that optimizes the encoding performance in terms of minimizing (3.12) or (3.13).

To distinguish it from $\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v,\ \sigma_w\right)$ (the average encoding performance in the case that the estimator only exploits statistical characterization of $w[n]$) we will use $\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v;\ \sigma_w\right)$ to denote the average encoding performance for a given set of σ_A, σ_v, and σ_w, where $w[n]$ is an IID Gaussian process of power level σ_w^2, and where $w[n]$ is known for estimation. For any given value of the random variable A, the Cramér-Rao bound on all unbiased estimates of A from $y[n]$, where $w[n]$ is known to the estimator and is a sample path of an IID Gaussian

[2]Although as $N \to \infty$ we can not rely on periodic inputs to achieve optimal performance, for any finite N, no matter how large, we can develop periodic inputs that are approximately optimal

process, satisfies

$$\mathcal{B}\left(A;\ y, p_w(\cdot)\right) = \left[\int \left\{\mathcal{B}\left(A-w;\ y\right)\right\}^{-1} p_w(w)\, dw\right]^{-1} \tag{3.28a}$$

$$= \sigma_v^2 \sigma_w \left[\int \frac{f^2\left(\frac{A-w}{\sigma_v}\right)}{Q\left(\frac{A-w}{\sigma_v}\right) Q\left(-\frac{A-w}{\sigma_v}\right)} f\left(\frac{w}{\sigma_w}\right) dw\right]^{-1} \tag{3.28b}$$

$$= \sigma_v \sigma_w \left[\int \frac{f^2(u)}{Q(u)\, Q(-u)} f\left(\frac{\sigma_v u + A}{\sigma_w}\right) du\right]^{-1} \tag{3.28c}$$

$$\approx \sigma_v \sigma_w f\left(\frac{A}{\sigma_w}\right) \mathcal{I} \tag{3.28d}$$

where approximation (3.28d) is valid for $\sigma_v \ll \sigma_w$, and where

$$\mathcal{I} = \left[\int_{\infty}^{\infty} \frac{f^2(u)}{Q(u)\, Q(-u)}\, du\right]^{-1} \approx 0.5536\ .$$

The average encoding performance $\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v;\ \sigma_w\right)$ is then given by substituting (3.28d) in (3.12)

$$\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v;\ \sigma_w\right) \approx \frac{\sigma_v \sigma_w \mathcal{I}}{\sigma_A} \int \frac{f\left(\frac{A}{\sigma_A}\right)}{f\left(\frac{A}{\sigma_w}\right)}\, dA \tag{3.29a}$$

$$= \frac{\sigma_v \sigma_w \mathcal{I}}{\sigma_A} \int_{\infty}^{\infty} \exp\left(-\frac{A^2}{2}\left[\frac{1}{\sigma_A^2} - \frac{1}{\sigma_w^2}\right]\right) dA \tag{3.29b}$$

which is finite if and only if $\sigma_w > \sigma_A$. In particular, the value of σ_w that minimizes $\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v;\ \sigma_w\right)$ for $\sigma_A \gg \sigma_v$ (*i.e.*, large SNR $\bar{\chi}$) is $\sigma_w = \sqrt{2}\,\sigma_A$, in which case (3.29b) reduces to

$$\bar{\mathcal{B}}\left(\sigma_A,\ \sigma_v;\ \sigma_w^{\text{opt}}\right) \approx \sqrt{8\pi}\,\mathcal{I}\,\sigma_v\,\sigma_A\ .$$

Specifically, if we select

$$\sigma_w^{\text{opt}}(\sigma_A) = \begin{cases} \sqrt{2\,\sigma_A^2 - \sigma_v^2} & \text{if } \sigma > \sqrt{2}\,\sigma_v \\ 0 & \text{otherwise} \end{cases}\ , \tag{3.30}$$

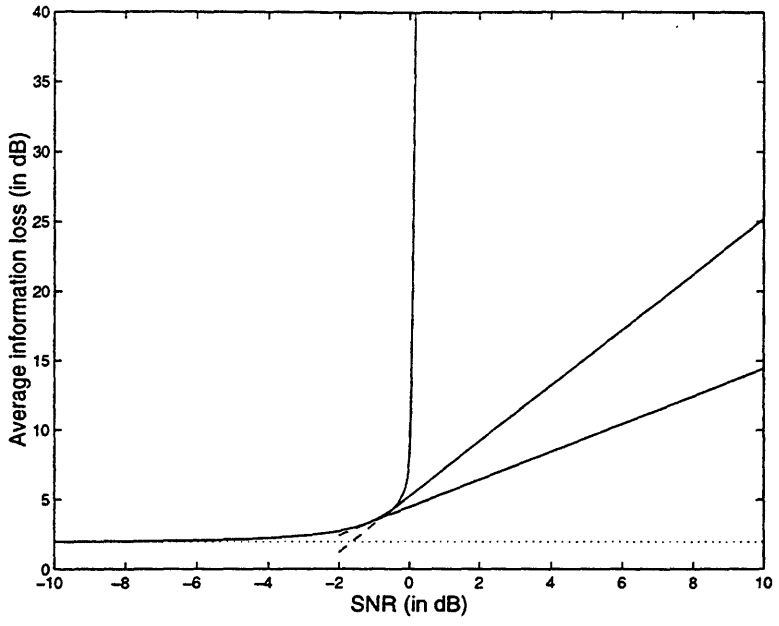


Figure 3-8: Average information loss as a function of signal-to-noise ratio $\bar{\chi}$ for no control inputs (upper solid) and for optimally designed pseudo-noise (middle solid) and known (lower solid) inputs in the case $M = 2$. Both the IID sensor noise and the *a priori* PDF are zero-mean Gaussian. The control input is a typical sample path of an IID Gaussian process of power level, selected according to (3.22) and (3.30), respectively. The successively lower dashed lines show the high-SNR performance, as predicted by (3.19) and (3.31), respectively. The dotted line depicts the 2 dB lower bound.

and by using the fact that $\bar{B}(A;\ s) = \sigma_v^2$, we get

$$\bar{\mathcal{L}}^{\mathbf{kn}}(\chi) \approx \sqrt{8\pi\mathcal{I}\,\bar{\chi}} \tag{3.31}$$

for high SNR, *i.e.*, by proper choice of the energy level of the Gaussian control input we can make the average information loss to grow as slow as linearly with SNR.

The information loss of this scheme is depicted in Fig. 3-8 as a function of SNR $\bar{\chi}$. The figure also depicts the high-SNR average performance for optimized pseudo-noise (upper dashed) and known (lower dashed) control inputs and predicted by (3.19) and the approximation (3.31), respectively. Although the encoding scheme and the criterion used to assess the encoding quality as well as the *a priori* assumptions about the information-bearing signal differ substantially from the unknown parameter case considered in Chapter 2, the resulting performance is strikingly similar.

The MAP estimator for any known control input sequence is very similar to (3.27) which

was used in the pseudo-noise case. Specifically, it can be readily implemented by means of the following EM algorithm which is a special case of the algorithm derived in App. A.5:

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathrm{m}_A + \frac{1}{1+\frac{\sigma_v^2}{N\sigma_A^2}}\left[\hat{A}_{\mathrm{EM}}^{(k)} - \mathrm{m}_A + \sum_{m=1}^{M} \frac{\sigma_v \mathcal{K}_{Y_m}\left(\mathbf{y}^N\right)}{\sqrt{2\pi}N} \frac{\exp\left(-\frac{\left(z_{m-1}^{(k)}\right)^2}{2}\right) - \exp\left(-\frac{\left(z_m^{(k)}\right)^2}{2}\right)}{Q\left(z_{m-1}^{(k)}\right) - Q\left(z_m^{(k)}\right)}\right], \tag{3.32a}$$

and where

$$z_m^{(k)} = \frac{X_m - \hat{A}_{\mathrm{EM}}^{(k)} - w_{\mathrm{opt}}}{\sigma_v}\ ; \tag{3.32b}$$

the MAP estimate $\hat{A}_{\mathrm{MAP}}\left(\mathbf{y}^N\right)$ is then given by

$$\hat{A}_{\mathrm{MAP}}\left(\mathbf{y}^N\right) = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}\ . \tag{3.32c}$$

Feedback in Control Input Selection

A priori information can also be incorporated in the system (2.1)–(2.2) employing feedback in the selection of the control sequence. Specifically, average encoding performance bounds and corresponding MAP estimators that asymptotically approach these bounds can be constructed.

Again we focus on the system corresponding to $M = 2$. We can design MAP estimators for the feedback case; these asymptotically (large N) attain the performance of the ML estimators based on feedback (developed in Chapter 2) and thus for any A asymptotically achieve the 2 dB bound. Consequently, in this case the average performance is the same.

In general, for finite N, the performance bounds of these MAP solutions will be dependent on N. In determining the lowest possible achievable Cramér-Rao bound for estimating $A \sim \mathcal{N}\left(\mathrm{m}_A, \sigma_A^2\right)$ based on observation of $\mathbf{y}^N$ from (2.1)–(2.2), we allow the selected control vector $\mathbf{w}^N$ to depend on the particular value of A. Specifically, let $\bar{B}\left(\mathbf{y}^N; \mathbf{w}^N\right)$ denote the Cramér-Rao bound for estimating A resulting from a particular selection method for the control input $\mathbf{w}^N$ based on observation of $\mathbf{y}^N$. We may use the Cramér-Rao bound on

unbiased estimates in the case that A is a random variable which in the case the input is known satisfies

$$\bar{\mathcal{B}}\left(\mathbf{y}^N;\, \mathbf{w}^N\right) = \left(E\left[\left\{\mathcal{B}\left(A;\, \mathbf{y}^N,\, \mathbf{w}^N\right)\right\}^{-1}\right] + \frac{1}{\sigma_A^2}\right)^{-1} \tag{3.33a}$$

$$= \left(\sum_{n=1}^{N} E\left[\left\{\mathcal{B}\left(A - w[n];\, y\right)\right\}^{-1}\right] + \frac{1}{\sigma_A^2}\right)^{-1} \tag{3.33b}$$

$$\geq \frac{1}{\frac{2N}{\pi\sigma_v^2} + \frac{1}{\sigma_A^2}}\,, \tag{3.33c}$$

where $\mathcal{B}\left(A;\, \mathbf{y}^N,\, \mathbf{w}^N\right)$ and $\mathcal{B}\left(A - w[n];\, y\right)$ are given by (2.23) and (2.16), respectively. Ineq. (3.33c) provides a bound on the performance of any unbiased estimator of A from $\mathbf{y}^N$, and for any selection of the control sequence $\mathbf{w}^N$. Note that (3.33c) results from application of (2.37), with equality achieved for $w[n] = -A$. Since such a control sequence is not plausible (due to its dependence on the unknown parameter A), in a manner analogous to (2.52) we may select the control sequence as follows

$$w[n] = -\hat{A}_{\mathrm{MAP}}\left(\mathbf{y}^{n-1}\right)\,. \tag{3.34}$$

The corresponding MAP estimator can be obtained from the ML estimation algorithm (2.53) with minor modifications, and can be derived as a special case of the algorithm described in App. A.5:

$$\hat{A}_{\mathrm{EM}}^{(k+1)}[n] = \mathrm{m}_A + \frac{1}{1 + \frac{\sigma_v^2}{n\sigma_A^2}}\left[\hat{A}_{\mathrm{EM}}^{(k)}[n] - \mathrm{m}_A + \sum_{m=1}^{n} \frac{\sigma_v\, y[m]}{\sqrt{2\pi n}} \frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{EM}}^{(k)}[n] - \hat{A}_{\mathrm{MAP}}[m]\right)^2}{2\sigma_v^2}\right)}{Q\left(\frac{\hat{A}_{\mathrm{EM}}^{(k)}[n] - \hat{A}_{\mathrm{MAP}}[m]}{\sigma_v}\, y[m]\right)}\right], \tag{3.35a}$$

$$\hat{A}_{\mathrm{MAP}}[n] \triangleq \hat{A}_{\mathrm{MAP}}\left(\mathbf{y}^n\right) = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}[n]\,. \tag{3.35b}$$

Empirical evidence suggests that the MAP estimate (3.35) in conjunction with selecting $w[n]$ according to (3.34) achieves the minimum possible information loss (3.33c) for mod-

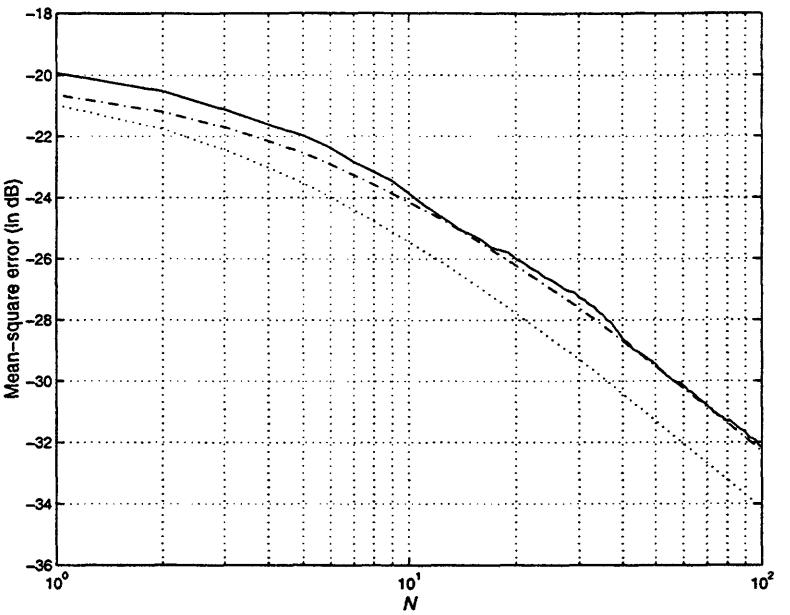


Figure 3-9: Performance based on Monte-Carlo simulations (solid curve) of the MAP estimator of the random parameter A based on observations from a binary quantizer where the control input at time n equals the negative of the estimate at time $n-1$. The dotted curves correspond to the Cramér-Rao bounds for estimating A based on the infinite-resolution sequence (dotted curve) and the quantized-sequence based on the best possible control sequence selection.

erate N values, similarly to its ML counterpart. Note that in the presence of a priori information, and for $\sigma_A \ll \sigma_v$, the control sequence $w[n]$ enables immediate operation around the quantizer threshold, and thus quicker convergence to the corresponding minimum possible information loss (3.33c). However, for large enough N, where the information from the available observations dominates the *a priori* information we may also substitute the MAP algorithm (3.35) with the low-complexity estimator (2.58) without compromising performance.

3.3 Unknown Noise Power Level

Another important extension of the estimation problem considered in Chapter 2 involves estimation of the unknown parameter of interest when in addition the noise power level is unknown. Specifically, consider the problem of estimating the unknown parameter A and

possibly the unknown noise power level σ_v from observation of

$$y[n] = F(A + \sigma_v\, \tilde{v}[n] + w[n]) \tag{3.36}$$

where $\tilde{v}[n]$ is an IID process of known statistical characterization, $w[n]$ is a control input, and $F(\cdot)$ is an M-level quantizer given by (2.2).

3.3.1 Performance Limits

In order to assess the performance of the encoding strategy we develop, we rely on extensions of the figures of merit developed in Chapter 2. Specifically, let $\boldsymbol{\theta} = [A\ \sigma_v]^T$ denote the vector of unknown parameters, and let also for convenience $\theta_1 = A$ and $\theta_2 = \sigma_v$. Let $\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)$ denote the 2×2 Cramér-Rao bound matrix for unbiased estimates of the vector parameter $\boldsymbol{\theta}$ from observation of $\mathbf{y}^N$. Then

$$E\left[\left(\hat{A}\left(\mathbf{y}^N\right) - A\right)^2\right] \geq \left[\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)\right]_{1,1},$$

and

$$E\left[\left(\hat{\sigma_v}\left(\mathbf{y}^N\right) - \sigma_v\right)^2\right] \geq \left[\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)\right]_{2,2},$$

where $\hat{A}\left(\mathbf{y}^N\right)$ and $\hat{\sigma_v}\left(\mathbf{y}^N\right)$ are any unbiased estimators of A and σ_v, respectively.

Analogously to the known σ_v case, we use as our measure of quality of the encoding strategy the following notion of information loss

$$\mathcal{L}(A,\, \sigma_v) = \frac{\left[\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)\right]_{1,1}}{\left[\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{s}^N\right)\right]_{1,1}}\,. \tag{3.37}$$

We assume that the range of the parameter of interest A is $(-\Delta,\, \Delta)$, while the unknown noise level satisfies $\sigma_{\min} < \sigma_v$. Worst-case performance is used as a measure of the encoding quality, *i.e.*,

$$\mathcal{L}_{\max}(\Delta,\, \sigma_{\min}) = \max_{\substack{A \in (-\Delta,\, \Delta) \\ \sigma_v \in (\sigma_{\min},\, \infty)}} \mathcal{L}(A,\, \sigma_v). \tag{3.38}$$

We focus our attention on the case where $\tilde{v}[n]$ is a zero-mean Gaussian noise process of unit

variance; similar results can be developed for nonGaussian admissible sensor noises.

Pseudo-noise Control Inputs

In this section we assume that the estimator can only exploit knowledge of the statistical characterization of the control input $w[n]$ for estimation. In particular, we assume that the control input can be modeled as an IID zero-mean Gaussian sequence of power σ_w^2. As usual, absence of a control input corresponds to the special case $\sigma_w = 0$. Consider the 2×2 Fisher Information matrix associated with the Cramér-Rao bound matrix $\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)$, *i.e.*,

$$\mathcal{F}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right) = \left[\mathcal{B}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)\right]^{-1} . \tag{3.39}$$

The $(i,\ j)$th entry of the Fisher Information matrix can be obtained by partial differentiation of the log-likelihood with respect to θ_i and θ_j, followed by an expectation, and can be put in the following form

$$\left[\mathcal{F}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)\right]_{i,j} = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{m=1}^{M} \dfrac{\gamma_m^2}{\epsilon_m} & \text{if } i = j = 1 \ (i.e.,\ \theta_i = \theta_j = A) \\ \dfrac{1}{N}\displaystyle\sum_{m=1}^{M} \dfrac{\delta_m^2}{\epsilon_m} & \text{if } i = j = 2 \ (i.e.,\ \theta_i = \theta_j = \sigma_v) \\ \dfrac{1}{N}\displaystyle\sum_{m=1}^{M} \dfrac{\gamma_m\,\delta_m}{\epsilon_m} & \text{if } i \neq j \end{cases} , \tag{3.40a}$$

where

$$\gamma_m = \frac{1}{\sigma_\alpha}\left[f\left(\frac{X_{m-1} - A}{\sigma_\alpha}\right) - f\left(\frac{X_m - A}{\sigma_\alpha}\right)\right] , \tag{3.40b}$$

$$\delta_m = \frac{\sigma_v}{\sigma_\alpha^2}\left[\frac{X_{m-1} - A}{\sigma_\alpha} f\left(\frac{X_{m-1} - A}{\sigma_\alpha}\right) - \frac{X_m - A}{\sigma_\alpha} f\left(\frac{X_m - A}{\sigma_\alpha}\right)\right] , \tag{3.40c}$$

$$\epsilon_m = Q\left(\frac{X_{m-1} - A}{\sigma_\alpha}\right) - Q\left(\frac{X_m - A}{\sigma_\alpha}\right) , \tag{3.40d}$$

$f(x) = \exp(-x^2/2)/\sqrt{2\,\pi}$, and $\sigma_\alpha = \sqrt{\sigma_v^2 + \sigma_w^2}$.

In the special case $M = 2$ the determinant of $\mathcal{F}\left(\boldsymbol{\theta};\ \mathbf{y}^N\right)$ equals zero, revealing that

estimation of $\boldsymbol{\theta}$ for $M = 2$ is an ill-posed problem. In the absence of pseudo-noise ($\sigma_w = 0$) this is easily clarified by noting that a parameter $\boldsymbol{\theta} = [A\ \sigma_v]^T$ yields the same observation sequence as the parameter $\lambda\,\boldsymbol{\theta} = [\lambda A\ \lambda\sigma_v]^T$; given any sequence $\tilde{v}[n]$, and by denoting the observed sequence as $y[n;\lambda]$ to denote its λ–dependence, we have

$$y[n;\lambda] = \mathrm{sgn}\,(\lambda A + \lambda\,\sigma_v\,\tilde{v}[n]) = \mathrm{sgn}\,(A + \sigma_v\,\tilde{v}[n]) = y[n;1]\,.$$

Similarly, in the pseudo-noise case, for any $\lambda > 1$, any pair $(A,\ \sigma_v)$ is equivalent to a pair $(\lambda A,\ \lambda'\sigma_v)$, where

$$\lambda' = \frac{1}{\sigma_v}\sqrt{\lambda^2\,(\sigma_v^2 + \sigma_w^2) - \sigma_w^2}$$

since

$$y[n;\lambda] = \mathrm{sgn}\,(\lambda A + \sqrt{\lambda'^2\,\sigma_v^2 + \sigma_w^2}\,\tilde{\alpha}[n]) = \mathrm{sgn}\,(\lambda A + \lambda\,\sigma_\alpha\,\tilde{\alpha}[n]) = y[n;1]\,.$$

For this reason, for pseudo-noise control inputs we focus on the case $M = 3$ to illustrate the encoder design. In particular, we assume that $F(\cdot)$ is a symmetric quantizer, *i.e.*, $X_1 = -X_2 = X$. Given Δ and $\sigma_{\min}$, we wish to select the noise power level σ_w so as to minimize the worst-case performance as depicted by (3.38).

The worst-case performance (3.38) in this case occurs at the parameter space boundary where $\sigma_v \to \sigma_{\min}$ and $|A| \to \Delta$. In particular, analogous to the case that the sensor noise power level is known we may define a measure of peak signal-to-noise ratio, as follows

$$\chi = \frac{\Delta}{\sigma_{\min}}$$

via which we can characterize the encoding performance. In Fig. 3-10 we show the optimal choice in terms of the pseudo-noise level in the sense of minimizing the worst-case information loss (3.38) as a function SNR for $\Delta = 1$ and $X = 0.5$. As in the case corresponding to a known sensor noise level examined in Chapter 2, it is evident that at high SNR $\Delta/\sigma_{\min}$, the optimal pseudo-noise level is independent of the sensor noise level $\sigma_{\min}$.

The solid line in Fig. 3-11 depicts the associated worst-case information loss as a function of SNR. In the same figure the dotted curve depicts the uncoded performance, corresponding

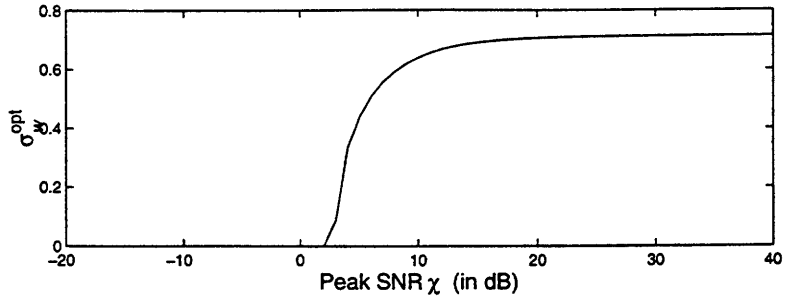


Figure 3-10: Optimal Pseudo-noise level as a function of SNR for a three level quantizer with $X = 0.5$, and $\Delta = 1$.

to $w[n] = 0$. For comparison we show the associated performance curves when there is no control input (dash-dot) and for pseudo-noise encoders (dashed) in the case that the sensor noise level is known (in which case the peak SNR equals Δ/σ_v).

Fig. 3-12 shows the additional information loss arising from lack of knowledge of the sensor noise level. As we can see, lack of knowledge of the sensor noise level comes at an additional cost of less than 8 dB encoding loss for any signal-to-noise ratio χ.

Known Control Inputs

We can also consider encoding strategies for the case that the estimator can fully exploit knowledge of the control input sequence used at the encoder. Naturally, we wish to construct the control input so as to minimize the worst-case information loss (3.38). In a fashion similar to the case where the noise level is known we can show that by using periodic control inputs of the form (2.26) where K is selected from (2.27) and where χ is replaced with $\Delta/\sigma_{\min}$ we can provide encoding strategies for which the associated information loss grows linearly with $\Delta/\sigma_{\min}$.

The (i, j)th entry of the Fisher Information matrix can be obtained by partial differentiation of the log-likelihood with respect to θ_i and θ_j, followed by an expectation:

$$\left[\mathcal{F}\left(\boldsymbol{\theta};\, \mathbf{y}^N\right)\right]_{i,j} = \begin{cases} \dfrac{K}{N}\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{K}\dfrac{\gamma_m^2[k]}{\epsilon_m[k]} & \text{if } i=j=1 \text{ (\textit{i.e.}, } \theta_i=\theta_j=A) \\ \dfrac{1}{N}\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{K}\dfrac{\delta_m^2[k]}{\epsilon_m[k]} & \text{if } i=j=2 \text{ (\textit{i.e.}, } \theta_i=\theta_j=\sigma_v) \\ \dfrac{1}{N}\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{K}\dfrac{\gamma_m[k]\,\delta_m[k]}{\epsilon_m[k]} & \text{if } i\neq j \end{cases} \quad , \quad (3.41a)$$

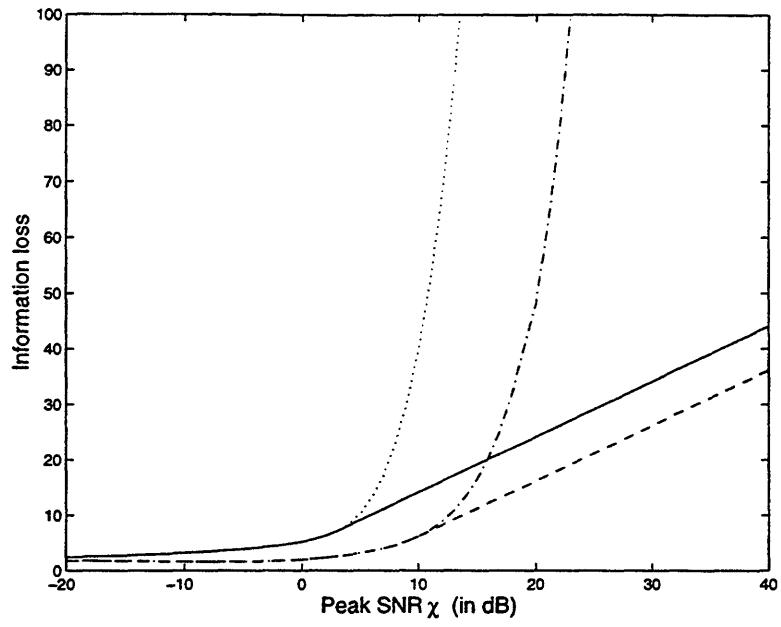


Figure 3-11: Information loss as a function of SNR in the absence of a control input (dotted) and in the presence of optimally selected pseudo-noise level (solid). For comparison, the associated performance curves for known sensor noise level are shown.

where

$$\gamma_m[k] = \frac{1}{\sigma_\alpha^2}\left[f\left(\frac{X_{m-1}-A-w[k]}{\sigma_\alpha}\right) - f\left(\frac{X_m-A-w[k]}{\sigma_\alpha}\right)\right], \tag{3.41b}$$

$$\delta_m[k] = \frac{\sigma_v}{\sigma_\alpha^2}\left[\frac{X_{m-1}-A-w[k]}{\sigma_\alpha} f\left(\frac{X_{m-1}-A-w[k]}{\sigma_\alpha}\right) - \frac{X_m-A-w[k]}{\sigma_\alpha} f\left(\frac{X_m-A-w[k]}{\sigma_\alpha}\right)\right], \tag{3.41c}$$

$$\epsilon_m[k] = Q\left(\frac{X_{m-1}-A-w[k]}{\sigma_\alpha}\right) - Q\left(\frac{X_m-A-w[k]}{\sigma_\alpha}\right), \tag{3.41d}$$

and $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$.

In Fig. (3-13) we show the performance (solid) in terms of the worst-case information loss as a function of SNR. We also show the associated performance in the case that the power level is known for estimation (dashed). As the figure illustrates, lack of knowledge of the noise power level comes at a cost that is upper-bounded by about 3 dB at low SNR, while at high SNR the additional loss is negligible.

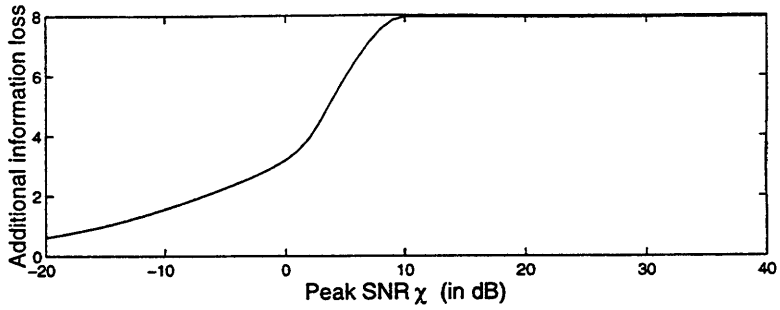


Figure 3-12: Additional worst-case information loss arising from lack of knowledge of the sensor noise level σ_v.

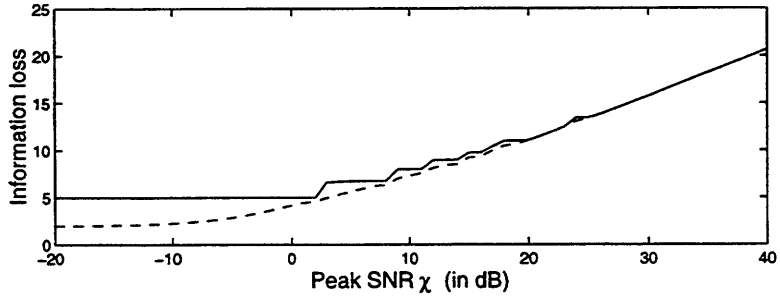


Figure 3-13: Worst-case information loss for known control input, in the case the sensor noise level is known (dashed) and unknown (solid).

Control Inputs in the Presence of Feedback

As in the known sensor noise level case, exploiting feedback in the design of the quantized encodings can yield substantial benefits in terms of the associated information and MSE loss. Although feedback can also be exploited in the case $M = 2$, for purposes of illustration we restrict our attention to the case $M = 3$ involving a symmetric quantizer. In that case, for any σ_v we have

$$\left[\mathcal{B}\left([A\ \sigma_v]^T;\ \mathbf{y}^N\right)\right]_{i,i} \leq \left[\mathcal{B}\left([0\ \sigma_v]^T;\ \mathbf{y}^N\right)\right]_{i,i}$$

for $i = 1, 2$, which reveals that

$$\left[\mathcal{B}\left([A\ \sigma_v]^T;\ \mathbf{y}^N,\ \mathbf{w}^N\right)\right]_{i,i} \geq \left[\mathcal{B}\left([0\ \sigma_v]^T;\ \mathbf{y}^N\right)\right]_{i,i}/N\,,$$

and where equality is achieved if $w[n] = -A$ for $n = 1, 2, \cdots, N$. In the presence of feedback the performance corresponding to $w[n] = -A$ can be practically achieved by using

encodings of the form:

$$w[n] = -\hat{A}[n-1] \tag{3.42}$$

where $\hat{A}[n-1]$ is a consistent estimate of A.

3.3.2 Estimation Algorithms

In App. B we present an EM algorithm which, under the condition that the likelihood function has a single local maximum over the parameter range of interest, results in the ML estimate of the unknown parameter vector $\boldsymbol{\theta} = [A \ \sigma_v]$. Depending on the particular case, this EM algorithm specializes to a number of different forms.

For pseudo-noise control inputs, the ML estimates $\hat{A}_{\mathrm{ML}}[N]$ and $\hat{\sigma}_{\mathrm{ML}}[N]$, of A and σ_v, respectively are given by

$$\hat{A}_{\mathrm{ML}}[N] = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)} \tag{3.43a}$$

$$\hat{\sigma}_{\mathrm{ML}}[N] = \lim_{k\to\infty} \sqrt{\left[\hat{\sigma}_{\mathrm{EM}}^{(k)}\right]^2 - \sigma_w^2}\,, \tag{3.43b}$$

where $\hat{A}_{\mathrm{EM}}^{(k+1)}$ and $\hat{\sigma}_{\mathrm{EM}}^{(k+1)}$ are given by (B.5a) and (B.5b) with $I = N$, and where

$$\underline{\sigma} = \sqrt{\sigma_{\min}^2 + \sigma_w^2} \tag{3.43c}$$

$$\overline{\sigma} = \infty \tag{3.43d}$$

$$B[k] = N\,\hat{A}_{\mathrm{EM}}^{(k)} + \sum_{m=1}^{M} \frac{\hat{\sigma}_{\mathrm{EM}}^{(k)}\,\mathcal{K}_{Y_m}\left(\mathbf{y}^N\right)}{\sqrt{2\,\pi}} \frac{\exp\left(\frac{\left(t_{m-1}^{(k)}\right)^2}{2}\right) - \exp\left(\frac{\left(t_m^{(k)}\right)^2}{2}\right)}{Q\left(t_{m-1}^{(k)}\right) - Q\left(z_m^{(k)}\right)} \tag{3.43e}$$

$$\frac{G[k]}{N} = \hat{A}_{\mathrm{EM}}^{(k)} + \left[\hat{\sigma}_{\mathrm{EM}}^{(k)}\right]^2 \left[1 + \sum_{m=1}^{M} \frac{\mathcal{K}_{Y_m}(\mathbf{y}^N)}{N\sqrt{2\,\pi}} \frac{\exp\left(\frac{\left(t_{m-1}^{(k)}\right)^2}{2}\right) t_{m-1}^{(k)} - \exp\left(\frac{\left(t_m^{(k)}\right)^2}{2}\right) t_m^{(k)}}{Q\left(t_{m-1}^{(k)}\right) - Q\left(z_m^{(k)}\right)}\right] \tag{3.43f}$$

$$t_m^{(k)} = \frac{X_m - \hat{A}_{\mathrm{EM}}^{(k)}}{\hat{\sigma}_{\mathrm{EM}}^{(k)}}\,. \tag{3.43g}$$

In Fig. 3-14 we present the MSE performance of this EM algorithm for $X = 0.5$, $N = 1000$, $A = 0.1$, $\Delta = 1$, $\sigma_{\min} = 0.05$, for several values of the pseudo-noise power level σ_v in the

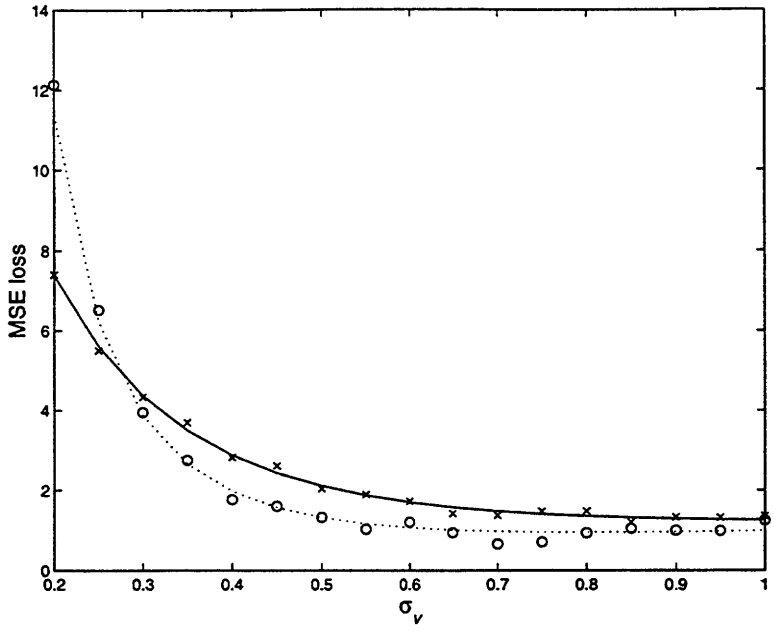


Figure 3-14: MSE loss in the parameter A from quantized encodings with pseudo-noise control inputs as a function of sensor noise level for $\sigma_w = 0.25$.

two cases that $\sigma_w = 0$ and $\sigma_w = 0.25$. As we can see, in both cases the information loss metric (3.37) accurately predicts the MSE loss performance of the EM algorithm.

Similarly, in Fig. 3-15 we depict the MSE in A and σ_v of the EM algorithm of App. B, when feedback is available and is exploited in the form of (3.42), for a symmetric quantizer with $M = 3$ and $X = 0.1$, in the case that $\boldsymbol{\theta} = [0.5\ 0.1]^T$. As the figure reveals, feedback in conjunction with the EM algorithm of App. B achieves the optimal performance within a few iterations.

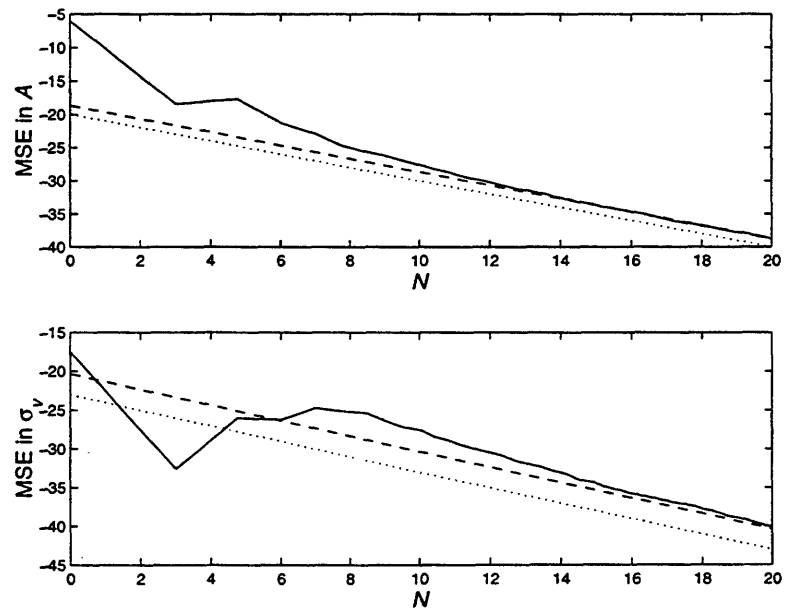


Figure 3-15: MSE performance of the EM algorithm of App. B for estimating the parameters A (upper figure) and σ_v (lower figure) from quantized encodings in the presence of feedback exploited via (3.42). The dashed lines correspond to the performance predicted by the Cramér-Rao bounds at $\boldsymbol{\theta}_* = (A_*, \sigma_v)$. The dotted lines correspond to the Cramér-Rao bounds for estimation of the parameter $\boldsymbol{\theta}$ based on original observations $\mathbf{s}^N$.

Chapter 4

Optimized Encoding Strategies for the Static Case

Encoders of the form of quantizer bias control are very attractive for digitally encoding noisy measurements since they can be designed to provide nice tradeoffs between encoder complexity and performance. However, although these encoders can achieve performance that does not degrade with SNR by exploiting feedback, these systems are inherently limited in the sense that, in general, they incur a small information loss.

In this chapter we examine the problem of eliminating performance losses by allowing more freedom in the encoder design. This problem may arise, for instance, in the context of distributed networks of wireless sensors, where bandwidth constraints limit the effective data rate (or equivalently bits per measurement) at which each sensor can reliably communicate to the host, but not the processing complexity at the sensor. The resulting problem of joint design of signal encoding and estimation can be viewed as a generalization of the low-complexity quantizer bias control systems developed in Chapter 2, where the encoder has the resources to perform more elaborate processing.

As in Chapters 2 and 3, we focus on the static case; we wish to determine the performance limits in terms of estimating a range-limited parameter based on digitally encoded noisy measurements. A block diagram description of the general problem is depicted in Fig. 4-1. A sequence of noise-corrupted observations $s[n]$ of an unknown parameter $A \in (-\Delta, \Delta)$ is encoded causally into a sequence of symbols M-ary symbols $y[n]$. The objective at the receiver is to estimate A based on the encodings $y[1], y[2], \cdots, y[n]$.

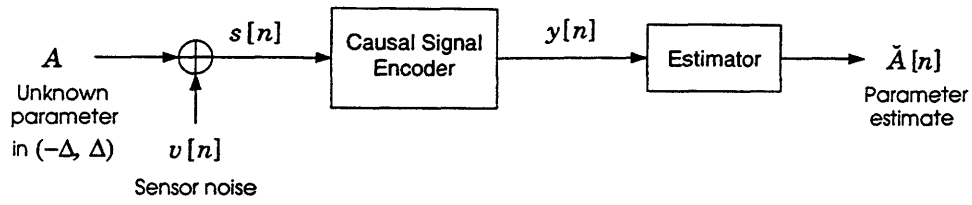


Figure 4-1: Block diagram of systems performing encoding and signal estimation.

We assume that system constraints actually limit the *average* encoding rate to at most one encoded M-ary symbol per sensor measurement (as shown in Fig. 4-1, this rate limitation is enforced by constraining the encoder to be causal). In the process we consider a variety of encoding schemes. These range from batch-mode encoding strategies, where the encoder first observes all n noisy measurements and then provides an n-symbol encoding from these measurements that can be used to form a single estimate at time n, to embedded fixed-rate encoders which encode one M-ary symbol per each available sensor measurement.

In Section 4.1 we introduce the figures of merit that we use to characterize the performance of the various encoding and estimator systems that we develop in this chapter. In Section 4.2 we develop variable-rate encoding methods and associated estimators which are asymptotically optimal in the sense that they asymptotically achieve the performance of any consistent estimator from the original sensor measurements that can be computed at the encoder. Then in Section 4.3 we consider fixed-rate encoding methods which encode at the sensor one symbol for every new available observation. We construct a class of such encoding methods which are also asymptotically optimal. We also illustrate the robustness of these encoding strategies by examining their performance in the presence of a nonadmissible noise. Finally, in Section 4.4 we present multi-sensor extensions of the single-sensor systems that are developed in this chapter.

4.1 Performance Characterization

For convenience, throughout this chapter we use the notation $\hat{A}[n]$ to denote an estimator of the parameter A that is formed at the sensor from the original noisy observations $s[1], s[2], \cdots, s[n]$ given by (2.3), and the notation $\check{A}[n]$ to denote an estimator of A that is formed at the host from all digitally encoded observations collected up to and including time

n. Throughout, we refer to $\hat{A}[n]$ and $\check{A}[n]$, as the sensor and the host estimate, respectively. Consistent with the system constraints, we design encoders for which at any time instant n the average encoding rate is less than or equal to one M-symbol per sensor measurement (*i.e.*, the number of encoded M-ary symbols never exceeds the number of available noisy measurements at the encoder).

In designing the signal encoder and estimator pair we use figures of merit based on the asymptotic performance of these systems. The performance metrics we employ are analogous to the asymptotic MSE loss criterion we have introduced in Chapter 2. Specifically, a naturally suited measure of performance is the asymptotic MSE loss, defined as

$$\mathcal{L}_{\text{MSE}}(A) \triangleq \lim_{n\to\infty} \frac{E\left[\left(\check{A}[n] - A\right)^2\right]}{\mathcal{B}\left(A;\, \mathbf{s}^n\right)} \,. \tag{4.1}$$

In all the encoding strategies we develop in this chapter, the encoder operates on a particular consistent sensor estimator $\hat{A}[n]$ formed from the original data $\mathbf{s}^n$. A suitable measure of encoding performance for these strategies is based on comparing the MSE of the host estimate $\check{A}[n]$ that is formed based on the encoding $\mathbf{y}^n$ against that of the associated estimate $\hat{A}[n]$ computed at the sensor from the original data. For that reason, we use the notion of asymptotic processing loss of an encoder with respect to a particular sensor estimator $\hat{A}[n]$ from $\mathbf{s}^n$, defined via

$$\mathcal{L}_{\text{Proc}}(A) \triangleq \lim_{n\to\infty} \frac{E\left[\left(\check{A}[n] - A\right)^2\right]}{E\left[\left(\hat{A}[n] - A\right)^2\right]} \,. \tag{4.2}$$

We refer to an encoding that achieves

$$\mathcal{L}_{\text{Proc}}(A) = 1 \,, \tag{4.3}$$

or

$$\mathcal{L}_{\text{MSE}}(A) = 1 \,, \tag{4.4}$$

over all $|A| < \Delta$ as asymptotically optimal, or asymptotically efficient, respectively. In the case that the sensor estimate $\hat{A}[n]$ (formed via $\mathbf{s}^n$) is asymptotically efficient with

respect to $\mathcal{B}(A;\ \mathbf{s}^n)$, the metrics $\mathcal{L}_{\text{MSE}}(A)$ and $\mathcal{L}_{\text{Proc}}(A)$ are identical and can be thus used interchangeably.

When designing an algorithm for encoding an asymptotically efficient sensor estimate $\hat{A}[n]$ formed from $\mathbf{s}^n$, it is important to minimize the mean-square difference between the sensor estimate $\hat{A}[n]$ and the associated host estimate $\check{A}[n]$, which we throughout refer to as the *residual error* between these two estimates. In particular, since the MSE of the sensor estimate $\hat{A}[n]$ decays as the inverse of the number of observations (for admissible noises), whenever we can design encoding schemes for which the residual error decays *faster* than the inverse of the number of observations, the resulting host estimate would be asymptotically optimal in the sense that it would satisfy (4.3). Specifically, if the residual error decays faster than $1/n$, *i.e.*, if

$$\lim_{n\to\infty} n\,E\left[\left(\hat{A}[n]-\check{A}[n]\right)^2\right]=0 \tag{4.5}$$

then, by using the triangle inequality

$$E\left[\left(\check{A}[n]-A\right)^2\right]\le E\left[\left(\hat{A}[n]-\check{A}[n]\right)^2\right]+E\left[\left(\hat{A}[n]-A\right)^2\right],$$

and the definition (4.2) we obtain

$$\mathcal{L}_{\text{Proc}}(A)\le 1+\lim_{n\to\infty}\frac{E\left[\left(\hat{A}[n]-\check{A}[n]\right)^2\right]}{E\left[\left(\hat{A}[n]-A\right)^2\right]}\le 1\,. \tag{4.6}$$

We also note that $1\le\mathcal{L}_{\text{Proc}}(A)$ due to the data processing inequality and the asymptotic efficiency of the sensor estimate, which, in conjunction with (4.6), proves the asymptotic optimality of the corresponding encoder in the sense of (4.3) and (4.4).

4.2 Variable-Rate Signal Encoders

In this section we consider algorithms that generate variable-rate encodings. Given any consistent sensor estimator $\hat{A}[n]$ formed from $\mathbf{s}^n$, the objective is to construct a digital encoding and a host estimator $\check{A}[n]$ from this encoding which (as a pair) asymptotically achieve the MSE performance of the original sensor estimator $\hat{A}[n]$. For reference, we first

consider asymptotically optimal batch-type algorithms, which collect all the available data prior to generating a digital encoding from which a host estimator $\check{A}[n]$ can be formed. Thereafter, we present a class of variable-rate algorithms which are extensions of the batch-mode algorithms and also result in asymptotically achieving the MSE rate of the original sensor estimator.

4.2.1 Batch-Type Encoders

The objective in batch-type encoding is to generate an encoded description that is to be used once at a particular instant n; a batch encoder first collects all the observations $s[1], s[2], \cdots, s[n]$ before forming the digital encoding $y[1], y[2], \cdots, y[n]$.

As suggested in the introduction of Chapter 2, it is straightforward to devise batch-type encoding algorithms and associated host estimators from these encodings that are asymptotically optimal, in the sense that they achieve (4.4). For convenience and without loss of generality, we consider the case where the encoder can construct an efficient sensor estimate $\hat{A}[n]$ from the noisy data $\mathbf{s}^n$, *i.e.*,

$$E\left[\left(A - \hat{A}[n]\right)^2\right] = \frac{\mathcal{B}(A;\ s)}{n}. \tag{4.7}$$

In that case, the encoder first collects $s[1], s[2], \cdots, s[n]$, subsequently computes $\hat{A}[n]$, and finally encodes as $y[1], y[2], \cdots, y[n]$ the n most significant M-ary symbols in the base-M representation of $\hat{A}[n]$. If the host estimate $\check{A}[n]$ of A based on the encoding $\mathbf{y}^n$ is formed as the real number whose base-M representation consists of the same n most significant M-ary symbols as the sensor estimate $\hat{A}[n]$ followed by some sequence of M-ary symbols, the residual error between $\hat{A}[n]$ and $\check{A}[n]$ decays to zero exponentially with n, *i.e.*,

$$E\left[\left(\check{A}[n] - \hat{A}[n]\right)^2\right] < D\,M^{-2n}, \tag{4.8}$$

and, hence, it satisfies (4.5). The constant D in (4.8) depends on the parameter range Δ and, in particular, satisfies $D \geq \Delta^2$. By using (4.7)–(4.8) and the triangle inequality, we obtain

$$\lim_{n\to\infty} n\,E\left[\left(\check{A}[n] - n\right)^2\right] = \mathcal{B}(A;\ s)\ ,$$

i.e., the host estimate $\check{A}[n]$ is asymptotically efficient with respect to $\mathcal{B}(A;\ s)$. In fact, $\check{A}[n]$ effectively achieves the bound $\mathcal{B}(A;\ s)/n$ rapidly with increasing n, since the residual error (4.8) between $\hat{A}[n]$ and $\check{A}[n]$ decays exponentially with n, while the MSE of the original sensor estimate $\hat{A}[n]$ in (4.7) decays only as $1/n$.

Similarly, if the estimate $\hat{A}[n]$ formed at the encoder via $\mathbf{s}^n$ is asymptotically efficient with respect to $\mathcal{B}(A;\ \mathbf{s}^n)$, the described batch encoding method produces an encoding from which the corresponding $\check{A}[n]$ is also asymptotically efficient. In general, if $\hat{A}[n]$ is any consistent estimator, the resulting $\check{A}[n]$ based on batch encoding is asymptotically optimal, in the sense that it satisfies (4.3). As we have discussed at the outset in Chapter 2, although this simple batch encoding method asymptotically achieves the MSE performance of the original sensor estimate $\hat{A}[n]$ (and the Cramér-Rao bound $\mathcal{B}(A;\ s)/n$ in case $\hat{A}[n]$ is asymptotically efficient), it has the major disadvantage that no encoded bits can be generated until all the observations are available. Moreover, it is not refinable, since no method is suggested for encoding any additional M-ary symbols as new sensor measurements are collected.

4.2.2 Refinable Variable-Rate Encoding Algorithms

As we have seen, batch encoding algorithms produce asymptotically optimal host estimates since the residual error between the sensor estimate and the host estimate decays at a faster rate than the mean-square error in the sensor estimate. In fact, the residual error between the two estimates decays exponentially fast with the number of observations, as opposed to the MSE in the encoder estimate which decays only as the inverse of the number of observations.

We can exploit this exponentially fast rate in the improvement of the host estimate quality to construct refinable variable-rate signal encoding strategies that are asymptotically optimal. In particular, by using repeatedly the batch-type encoding algorithm at a sequence of appropriately spaced time instants N_k, we can construct variable-rate encoding strategies which achieve (4.3) and for which the average encoding rate never exceeds one M-ary symbol per observation. Specifically, at each $n = N_k$ for $k = 1, 2, \cdots$ we may use the batch-type algorithm to encode the sensor estimate $\hat{A}[N_k]$ obtained from $\mathbf{s}^{N_k}$ into the $(N_k - N_{k-1})$ M-ary symbols $y[N_{k-1}+1], \cdots, y[N_k-1], y[N_k]$, based on which the host estimate $\check{A}[N_k]$ is to be formed. Since no encoded symbols are supplied by the encoder to the host between

time instants $n = N_{k-1} + 1$ and $n = N_k - 1$, the host may use as its estimate for all these time instants the most current host estimate, namely, $\check{A}[N_k - 1]$.

Note that N_{k-1} and N_k must be spaced far enough from one another so that the number of M-ary symbols used to describe the sensor estimate $\hat{A}[N_k]$ (*i.e.*, $N_k - N_{k-1}$) is large enough to guarantee that the residual error decays faster than $1/N_k$. On the other hand, since no encoded symbols are to be communicated between $n = N_{k-1} + 1$ and $n = N_k - 1$, the time instances N_{k-1} and N_k should still be close enough so that during the delay incurred by this batch-type scheme the "old" host estimate $\check{A}[N_k - 1]$ remains still "accurate enough". The following theorem describes how these instants N_k can be spaced in time so as to guarantee that the residual error between the host estimate $\check{A}[n]$ and the sensor estimate $\hat{A}[n]$ decays faster than $1/n$.

Theorem 1 *Let*

$$N_{k+1} = N_k + h[N_k] \tag{4.9}$$

for $k \geq 1$, initialized with $N_1 \geq 1$, and where $h : \mathbf{N}^+ \to \mathbf{N}^+$. Consider the encoding strategy which at time $n = N_{k+1}$ encodes as $y[N_k + 1], y[N_k + 2], \cdots, y[N_{k+1}]$ the $h[N_k]$ most significant symbols in the base-M representation of a consistent sensor estimator $\hat{A}[n]$ from $\mathbf{s}^n$. Let $\bar{A}[N_{k+1}]$ denote the number whose $h[N_k]$ most significant symbols in the base-M representation is given by $y[N_k+1], y[N_k+2], \cdots, y[N_{k+1}]$ followed by 0's. If the function $h[\cdot]$ satisfies both

$$\lim_{n\to\infty} \frac{h[n]}{n} = 0\,, \tag{4.10a}$$

and

$$\limsup_{n\to\infty} \frac{\ln(n)}{h[n]} < 2\,\ln(M)\,, \tag{4.10b}$$

then the host estimator $\check{A}[n]$ given by

$$\check{A}[n] = \bar{A}[\max_{k:N_k \leq n} N_k]\,, \tag{4.11}$$

is asymptotically optimal in the sense that it achieves (4.3).

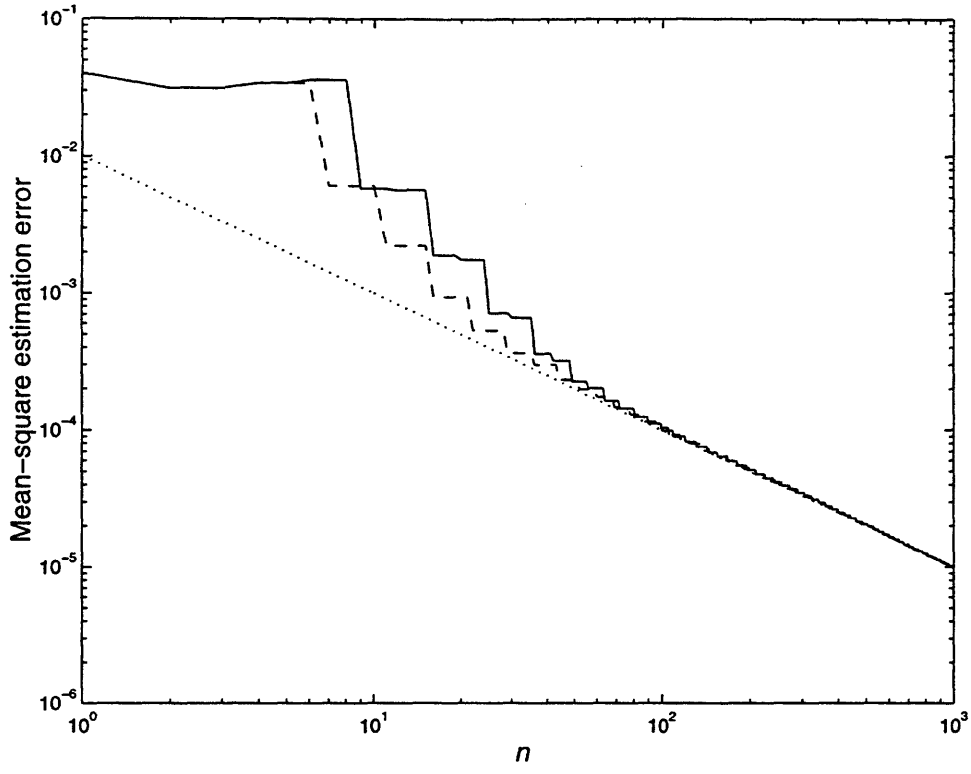


Figure 4-2: MSE performance of $\check{A}[n]$ from (4.11) in Gaussian noise, where $\hat{A}[n]$ is the sample mean (4.12). Simulation parameters: $\Delta = 1$, $A = 0.2$, $\sigma_v = 0.1$.

A proof of the asymptotic optimality of the encoding class of Theorem 1 is included in App. C.1.

Fig. 4-2 depicts the MSE performance of two instances of the host estimator $\check{A}[n]$ from (4.11) in the case that $M = 2$, $v[k] \sim \mathcal{N}(0, \sigma_v^2)$, and where the estimator based on $\mathbf{s}^n$ is the sample-mean, *i.e.*,

$$\hat{A}[n] = \frac{1}{n} \sum_{k=1}^{n} s[k] \, . \tag{4.12}$$

In particular, the solid and the dashed curves in the figure depict the MSE performance of $\check{A}[n]$ in the two special cases that the function $h[\cdot]$ in (4.11) is given by

$$h[n] = \lceil \sqrt{n} \rceil \, , \tag{4.13}$$

and

$$h[n] = \left\lceil 2\sqrt{\ln(n)} \right\rceil , \tag{4.14}$$

respectively, and where $\lceil x \rceil$ denotes the smallest integer that is greater or equal to x. In both cases the recursion (4.11) is initialized with $N_1 = 2$. One can easily verify that both (4.13) and (4.14) satisfy (4.10). The dotted line in the figure depicts the Cramér-Rao bound $\mathcal{B}(A;\ \mathrm{s}^n)$ which in this Gaussian case is achieved by the sample-mean estimator (4.12) for all n. As the figure illustrates, both host estimates are asymptotically efficient with respect to this bound. Clearly, the particular choice of N_1 and $h[\cdot]$ dictates how fast the MSE of the host estimate $\check{A}[n]$ achieves the bound $\mathcal{B}(A;\ \mathrm{s})^n$. In general, optimal selection of N_1 and $h[\cdot]$ depends on the signal-to-noise ratio Δ/σ_v, and the particular sensor noise and sensor estimate characteristics.

4.3 Fixed-Rate Encodings

Although the host estimate $\check{A}[n]$ in (4.11) is optimal in the sense that it asymptotically achieves the MSE rate of the sensor estimate $\hat{A}[n]$ formed from the original original measurements, the encoded sequence is not generated at a fixed encoding rate. In particular, delay is inherently incurred by the encoder and this encoding delay increases with n. In general, we may want to construct embedded algorithms that generate fixed-rate data encoding descriptions, *i.e.*, algorithms which provide one M-ary symbol of the description for each new available observation.

This problem possesses many similarities to the one of successive refinement of information [17]. In that problem a sequence of n IID random variables of known PDF are observed, and the task is to form a successively refinable approximate description of these n observations which achieves optimal or close to optimal approximation quality at any description level, as measured by a rate distortion metric. Analogously, in the problem we are addressing in this section the PDF of the IID random variables $s[n]$ is known up to an uncertainty in the mean. The task is to form a multi-stage encoding, so that the estimate sequence $\check{A}[n]$ generated from the nth stage encoding $\mathbf{y}^n$ is asymptotically optimal, *i.e.*, the MSE performance of $\check{A}[n]$ achieves $\mathcal{B}(A;\ \mathrm{s}^n)$ for any n large enough.

In this section we develop fixed-rate digital encodings that result in asymptotically

efficient estimation with respect to $\mathcal{B}(A;\ s)$. We focus on the case $M = 2$, although similar asymptotically optimal schemes can also be designed for $M > 2$[1].

In general, an embedded fixed-rate binary encoder is a rule for selecting the nth encoded bit $y[n]$ based on $\mathbf{y}^{n-1}$ and $\mathbf{s}^n$. The approach we follow in this section constrains the encoding strategy to select $y[n]$ based on $\mathbf{y}^{n-1}$ and $\hat{A}[n]$, where $\hat{A}[n]$ denotes a rule for obtaining an estimate of A based on $\mathbf{s}^n$. In that sense the objective of the encoding strategy is to achieve the performance of a particular estimator $\hat{A}[n]$ from the original observations.

The motivate the design of the encoders we present in this section, it is worthwhile to revisit the encoders we have developed in Chapter 2, which also generate fixed-rate digitally encoded descriptions. As we have seen, encodings in the form of quantizer bias control via feedback such as (2.59), when used in conjunction with the associated host estimators developed in Section 2.3.3, come within $10\log_{10}\mathcal{L}(A_*)$ dB of the optimal performance $\mathcal{B}(A;\ s)$. For instance, in the Gaussian scenario, selecting

$$y[n] = \operatorname{sgn}\left(s[n] - \check{A}[n-1]\right) \tag{4.15}$$

with $\check{A}[n-1]$ given by the linear-complexity algorithm (2.58) results in a 2 dB loss.

A lot of insight can be gained in the design of the asymptotically optimal schemes by examining the performance limits of the low-complexity structure (2.58), originally developed for signal estimation from the encodings generated by (4.15). Specifically, it is instructive to consider the MSE performance of the host estimate $\check{A}[n]$ given by the low-complexity algorithm (2.58), in the case that the sensor *precisely* knows the static signal A, in which case it could use an encoding of the form (4.15) where the noisy measurements $s[n]$ are replaced by A. The MSE performance of the resulting host estimate $\check{A}[n]$ is described by the following theorem which we prove in App. C.2.

Theorem 2 *Given $0 < c < \infty$, consider the dynamical system*

$$\check{A}[n] = \check{A}[n-1] + \frac{c}{n}\operatorname{sgn}\left(A - \check{A}[n-1]\right)\ , \tag{4.16}$$

[1]In fact, we can easily design asymptotically efficient schemes for $M > 2$, by trivial extensions of the $M = 2$ case, namely, by exploiting at the host only two of the available M encoded levels. Although beyond the scope of this thesis, designing algorithms that generate optimized fixed-rate M-level encoded schemes is clearly a problem worth further investigation.

initialized with $\check{A}[n_o]$, *for some* $n_o \geq 1$. *Then*

$$\lim_{n\to\infty} \check{A}[n] = A \tag{4.17}$$

for any n_o, *any initialization* $\check{A}[n_o]$, *and any* A. *In addition, the mean-square difference between* A *and* $\check{A}[n]$ *decays as* $1/n^2$. *In particular, for almost all initial conditions*

$$\limsup_{n\to\infty}\ n^2\,|\check{A}[n] - A|^2 = c^2\ . \tag{4.18}$$

As suggested in Theorem 2, in the case that an error-free estimate is available and used for encoding at the sensor the residual error decays as $1/n^2$. In the actual setting where noisy measurements of A are instead available at the sensor, replacing $s[n]$ with a sensor estimate $\hat{A}[n]$ can actually improve the MSE performance of the host estimate. In particular, we consider the following binary encoding method

$$y[n] = \mathrm{sgn}\left(\hat{A}[n] - \check{A}[n-1]\right), \tag{4.19a}$$

where the host estimate $\check{A}[n]$ based on $\mathbf{y}^n$ is given by

$$\check{A}[n] = \begin{cases} \text{from look-up table} & \text{if } n \leq n_o \\ \mathcal{I}_\Delta\left(\check{A}[n-1] + \frac{\lambda\sigma_v}{n}\, y[n]\right) & \text{if } n > n_o \end{cases}, \tag{4.19b}$$

and where we are also interested in optimally selecting the parameter λ. As in Chapter 2, we assume that $v[n] = \sigma_v\,\tilde{v}[n]$; in that sense optimal selection of λ will depend on $p_{\tilde{v}}(\cdot)$, the particular sensor noise generating PDF[2]. The estimator that is to be used with this encoder is given by (4.19b). Note the similarity of algorithms (4.19b) and (2.58) in terms of their dependence on the sensor noise scaling σ_v.

The look-up table structure in (4.19b) is used to provide faster convergence to asymptotically optimal behavior. Specifically, at high peak SNR (*i.e.*, for $\Delta \gg \sigma_v$), the first collection of bits $y[1], \cdots, y[n_o]$ may be used to encode a coarse description of A, ignoring

[2]More generally, we may consider nonstationary λ's, *i.e.*, λ's of the form $\lambda[n] = f(\mathbf{y}^n)$. Although proper design of the resulting encoders can potentially further improve the performance of the systems presented in this section, its investigation is beyond the scope of this thesis.

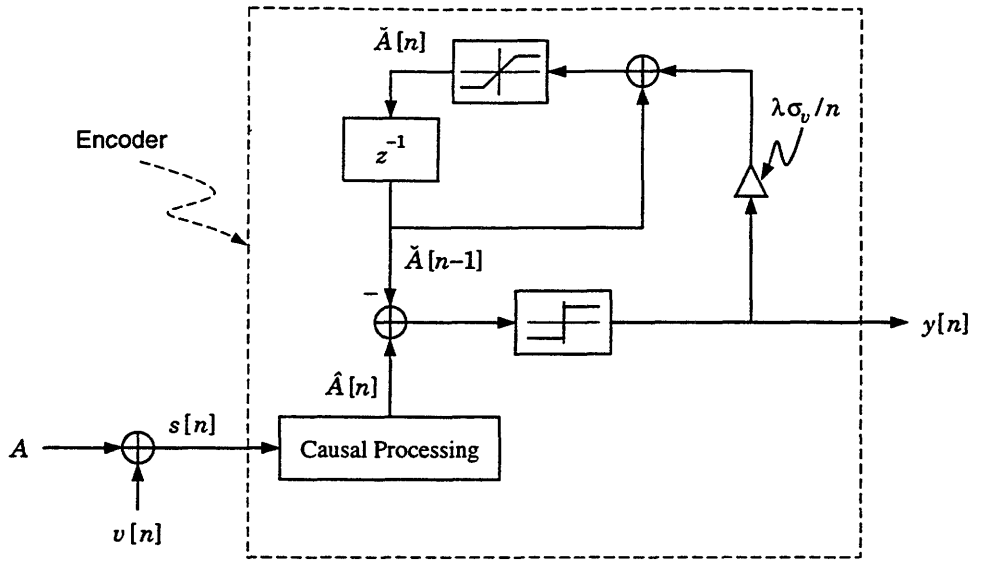


Figure 4-3: Block diagram of the sequential encoder (4.19) for $n > n_o$.

the effects of sensor noise, *i.e.*,

$$\check{A}[n] = \check{A}[n-1] + \frac{\Delta}{2^n}\, y[n] \quad \text{for } n \leq n_o \,, \tag{4.20}$$

initialized with $\check{A}[0] = 0$, and where $y[n]$ is given by (4.19a). Naturally, selection of a suitable value for n_o in (4.19b) depends on Δ, σ_v, and the particular sensor noise PDF. We must note, however, that the proper value of n_o depends logarithmically on $\chi = \Delta/\sigma$, as (4.20) suggests about the convergence of the MSE of $\check{A}[n]$ to that of $\hat{A}[n]$ (see also the associated discussion in Section 2.3.3).

Since we are primarily concerned with the performance of the algorithm (4.19) for large n, we focus on the encoding performance for $n > n_o$. The block diagram for the sequential encoder of observations $s[n]$ into bits $y[n]$ for $n > n_o$ is shown in Fig. 4-3. Intuitively, at any time instant n, the sensor encodes the sign of the difference between the current sensor estimate $\hat{A}[n]$ and the most recent host estimate $\check{A}[n-1]$. The associated host decoder/estimator from these bits is also shown in Fig. 4-4. As we have remarked, for $n < n_o$ both the decoder and encoder may employ a lookup table to obtain $\check{A}[n]$.

This class of fixed-rate binary encoding and estimator pairs have very attractive asymptotic properties. In particular, we may recall that (4.18) provides a bound on the best possible decay rate of the mean-square difference between $\hat{A}[n]$ and $\check{A}[n]$ of any algorithm of the form (4.19), since the algorithm (4.19) encodes an estimate $\hat{A}[n]$ rather than A which

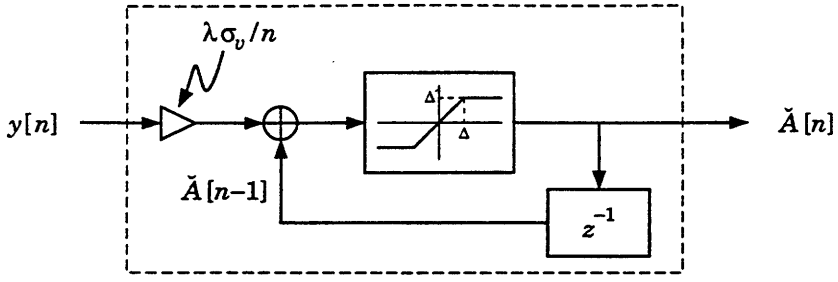


Figure 4-4: Block diagram of the sequential decoder associated with the encoder described by (4.19) for $n > n_o$.

was used in Theorem 2 (see (4.16)). As we will see in the following sections, if the MSE of $\hat{A}[n]$ decays as $1/n$ and if $\hat{A}[n]$ satisfies a mild set of conditions, the resulting residual error between the host and the sensor estimate decays as $1/n^2$, guaranteeing that the resulting host estimate $\check{A}[n]$ is asymptotically optimal, in the sense that it achieves (4.3).

4.3.1 Gaussian Sensor Noise

For Gaussian sensor noise, fixed-rate encodings of the form (4.19) can be designed that possess low computational complexity and are asymptotically efficient in the sense that they achieve (4.4). Specifically, consider the encodings of the form (4.19) where the sensor estimate $\hat{A}[n]$ is the sample-mean of the original noisy measurements, given by (4.12). Then, for sufficiently large n we can easily show that

$$E\left[\left(\hat{A}[n]-\check{A}[n-1]\right)^2\right] \approx \beta\,\sigma_v^2/n^2 \tag{4.21}$$

for some $\beta > 1$. In particular, as shown in App. C.3,

$$\lim_{n\to\infty}\; n^2\,E\left[\left(\hat{A}[n]-\check{A}[n-1]\right)^2\right] = \beta\,\sigma_v^2\,. \tag{4.22}$$

Using the readily verified identity satisfied by the sample-mean estimate

$$E\left[\left(\hat{A}[n]-\hat{A}[n-1]\right)^2\right] = \frac{\sigma_v^2}{n^2}\left(1+\frac{1}{n+1}\right), \tag{4.23}$$

the triangle inequality, and (4.22) we obtain the following bounds on the residual error

$$(\beta-1)\,\sigma_v^2 \le \lim_{n\to\infty}\; n^2\,E\left[\left(\hat{A}[n-1]-\check{A}[n-1]\right)^2\right] \le (\beta+1)\,\sigma_v^2. \tag{4.24}$$

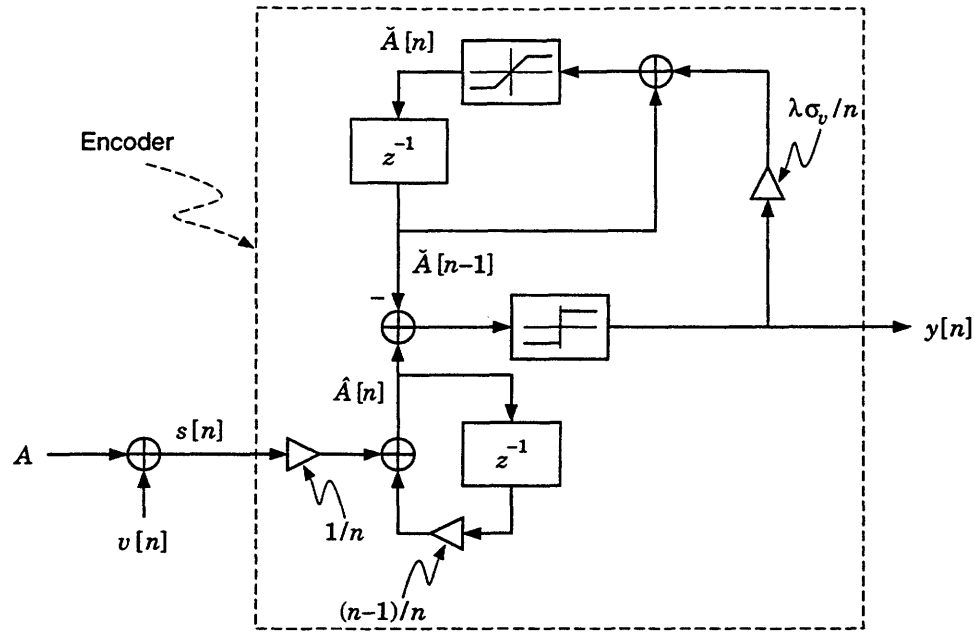


Figure 4-5: Block diagram of the sequential encoder (4.12)–(4.19b) for $n > n_o$, for asymptotically efficient estimation in white Gaussian noise.

The set of Ineqs. (4.24) implies that the residual between $\hat{A}[n]$ and $\check{A}[n]$ decays as $1/n^2$. Hence, since the sample-mean $\hat{A}[n]$ is an efficient estimator in additive IID Gaussian noise, we have

$$\begin{aligned} \mathcal{B}(A;\ s) \leq \lim_{n\to\infty} n\, E\left[\left(\check{A}[n]-A\right)^2\right] &\leq \lim_{n\to\infty} n\left\{E\left[\left(\check{A}[n]-\hat{A}[n]\right)^2\right]+E\left[\left(\hat{A}[n]-A\right)^2\right]\right\} \\ &\leq \mathcal{B}(A;\ s)\ , \end{aligned}$$

which reveals that the host estimate $\check{A}[n]$ formed from $\mathbf{y}^n$ is asymptotically efficient with respect to $\mathcal{B}(A;\ \mathbf{s}^n)$.

In this case, the block diagram of the sequential encoder of the original noisy observations $s[n]$ into bits $y[n]$ for $n > n_o$ specializes to the low-complexity structure shown in Fig. 4-5. The associated decoder/estimator from these bits is also shown in Fig. 4-4. For $n < n_o$ both the decoder and encoder may obtain $\check{A}[n]$ by means of the same lookup table. Again, in the coarse stage of the description (*i.e.*, $n < n_o$) the residual error between the two estimates decays exponentially with n since the noise level is small compared to the dynamic range.

Although the host estimate $\check{A}[n]$ obtained via (4.19) and (4.12) is asymptotically efficient for any $\lambda > 0$ in (4.19b), we can choose the value of λ so as to minimize the residual error given by (4.21). Specifically, as we show in App. C.3, β in (4.21) and λ in (4.19b) are related

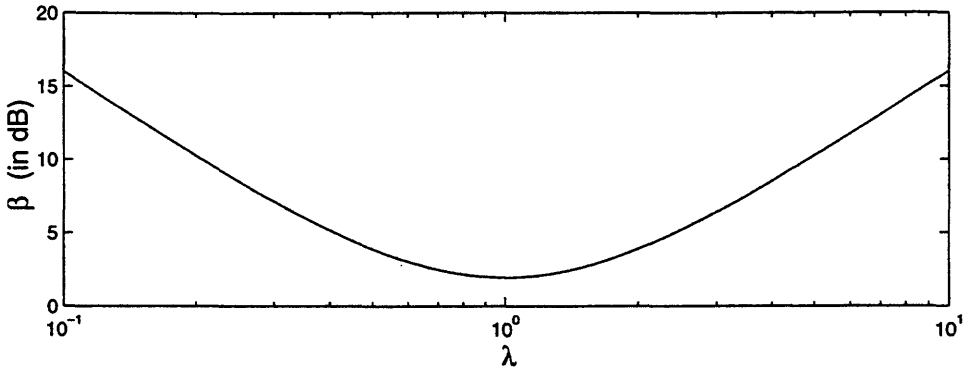


Figure 4-6: Resulting residual error scaling β as a function of parameter λ in (4.19b).

as follows

$$\beta(\lambda) = \frac{\pi \left(1+\lambda^2\right)^2}{8\,\lambda^2}, \tag{4.25}$$

so that

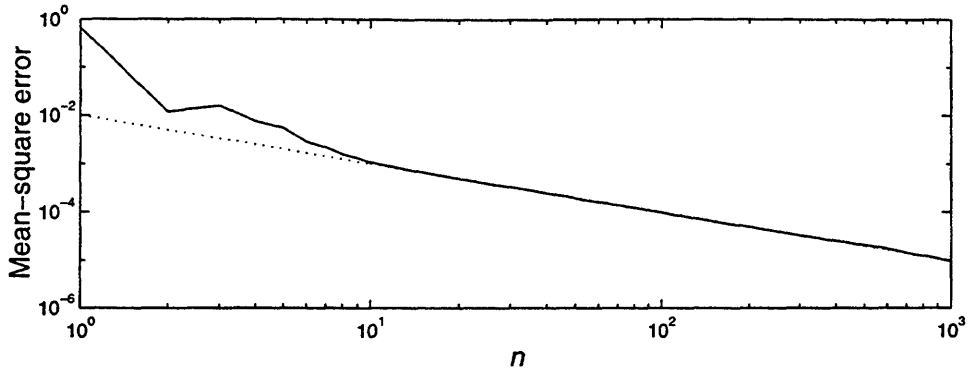
$$\min_{\lambda} \beta(\lambda) = \pi/2,$$

which is achieved for $\lambda = 1$. Fig. 4-6 depicts the residual error dependency on the value of λ chosen. As (4.25) reveals, selecting a value of λ an order of magnitude larger or smaller that the optimal value results in increasing the residual error by about 14 dB for any given n.

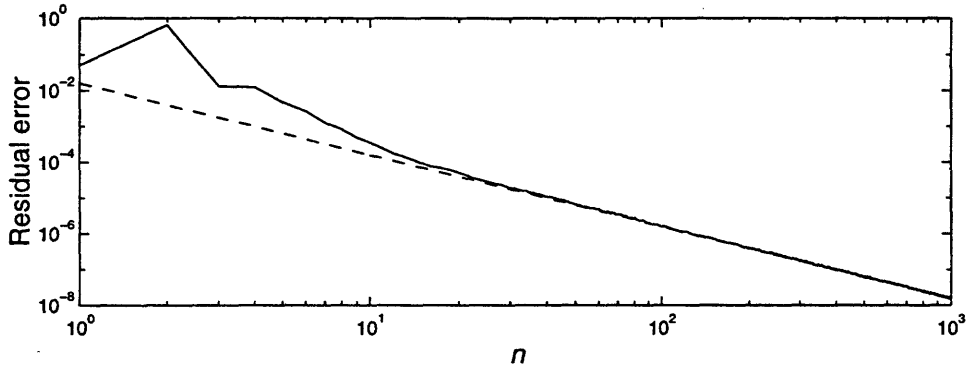
Fig. 4-7 illustrates the validity of our analysis for this Gaussian scenario by means of Monte-Carlo simulations for a specific example where $\Delta = 1$, $\sigma_v = 1$, $A = 0.2$, and $\lambda = 1$. The dotted line in Fig. 4-7(a) represents the Cramér-Rao bound $\mathcal{B}(A;\, \mathbf{s}^n)$ while the solid curve depicts the MSE in the host estimate $\check{A}[n]$ as a result of Monte-Carlo simulations. Fig. 4-7(b) depicts the associated residual error (4.21) from simulations (solid) and the estimate of the residual error (dashed) obtained via (4.21) and (4.25).

4.3.2 Robust Encodings in NonGaussian Finite-Variance Noise

The low-complexity encoding method consisting of (4.19) and the sample-mean $\hat{A}[n]$ in (4.12) is very robust with respect to variations in the sensor noise PDF. In particular, it achieves similar performance characteristics when the sensor noise $v[n]$ is an IID finite-



(a) Mean-square estimation error in $\check{A}[n]$



(b) Residual error

Figure 4-7: Performance of $\check{A}[n]$ from (4.19b), where $y[n]$ is given by (4.19a) and $\hat{A}[n]$ is the sample mean (4.12).

variance nonGaussian admissible process, as we now show.

For convenience, we assume that $\tilde{v}[n]$ is a unit-variance distribution, in which case σ_v^2 equals the variance of the sensor noise $v[n]$. Without loss of generality we consider the case where the sensor noise is zero-mean, in which case, as is well known, the sample-mean $\hat{A}[n]$ from (4.12) forms a consistent estimator of A with MSE equal to σ_v^2/n. The method that we used in App. C.3 to show that the host estimate $\check{A}[n]$ of the previous section has asymptotic MSE equal to σ_v^2/n applies exactly to this case as well; that is, the encoder/estimator structure described by (4.12) and (4.19) provides a host estimate $\check{A}[n]$ with asymptotic MSE equal to the noise power level divided by the available number of observations. Conveniently, the encoder and the decoder do not even require knowledge of

σ_v to obtain an estimate; it is simply required that both the encoder and decoder use the same system parameters, *i.e.*, the same lookup table and value of $\lambda\,\sigma_v$ in (4.19b). However, knowledge of σ_v can be exploited to provide faster convergence to asymptotic efficiency via optimal selection of $\lambda\,\sigma_v$.

Although attractive due to its simplicity, this approach does not always lead to a better asymptotic MSE than the quantizer bias control encoding approach described by (4.15) and (4.19b). This is clearly illustrated in the special case where the sensor noise is Laplacian, *i.e.*,

$$p_v\,(v) = \frac{1}{\sqrt{2}\,\sigma_v}\,e^{-\sqrt{2}\,|v|/\sigma_v}\ .$$

The Cramér-Rao bound for estimating A from $s[n]$ can be obtained by partial differentiation of the log-likelihood function followed by an expectation and is given by

$$\mathcal{B}\,(A;\ \mathbf{s}^n) = \frac{\sigma_v^2}{2\,n}. \tag{4.26}$$

Note that in this case the sample-mean $\hat{A}[n]$ is not asymptotically efficient with respect to $\mathcal{B}\,(A;\ \mathbf{s}^n)$ from (4.26), since it incurs a $10\log_{10}2 \approx 3$ dB loss. Hence, the encoder/estimator structure (4.19) that operates on the sample-mean $\hat{A}[n]$ incurs a 3 dB loss. Alternatively, consider using the quantizer bias control-based encoder/estimator structure described by (4.15) and (4.19b). The associated information loss (2.4) in the case of the Laplacian PDF is minimized at $A_* = 0$; by using the expression for $\mathcal{B}\,(A;\ \mathbf{y}^n)$ given by (2.15) with $\alpha = v$, we obtain

$$\mathcal{L}(A_*) = \frac{\mathcal{B}\,(0;\ y)}{\sigma_v^2/2} = 1\,.$$

Interestingly, at $A_* = 0$ this quantizer bias control encoder with feedback incurs *no* information loss. Hence, in the Laplacian case we may expect the quantizer bias control-based method described by (4.15) and (4.19b) to asymptotically outperform the encoding/estimation set (4.12)–(4.19b).

This is indeed the case as demonstrated in Fig. 4-8 where we depict the MSE performance of these two methods in the Laplacian case, for $\Delta = 1$, $\sigma_v = 0.1$, along with $\mathcal{B}\,(A;\ \mathbf{s}^n)$ (lower dotted line) and the MSE of the sample mean (upper dotted line). As we can see,

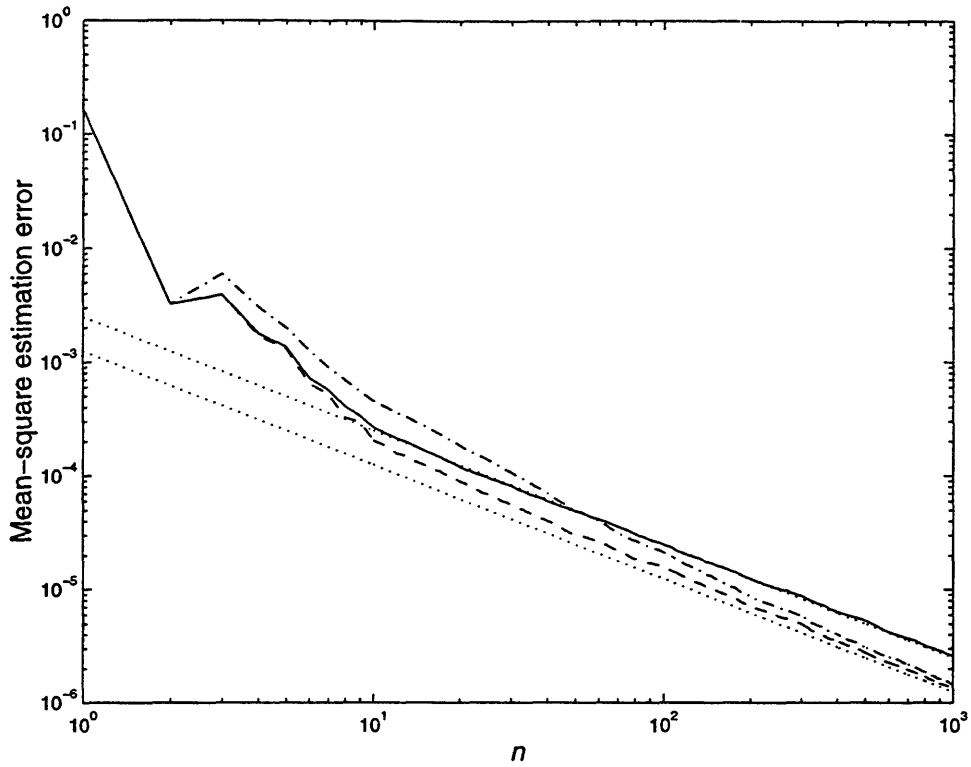


Figure 4-8: MSE performance of the host estimator in Laplacian sensor noise. The sensor estimate encoded in each case is the sample-mean (solid), the sensor measurement $s[n]$ (dash-dot), and the ML estimate (dashed). The two dotted lines depict the Cramér-Rao bound for estimating A given $\mathbf{s}^n$ (lower) and σ_v^2/n (upper).

the method encoding the difference between the sample mean and the current host estimate (solid curve) asymptotically achieves the sample-mean MSE rate (*i.e.*, a 3 dB loss), whereas the quantizer bias control method encoding the difference between $s[n]$ and $\check{A}[n-1]$ (dash-dot curve) leads to an estimate that is asymptotically efficient with respect to the original observation sequence $s[n]$.

4.3.3 NonGaussian Admissible Noise

Whenever the sensor can form an estimate $\hat{A}[n]$ from $\mathbf{s}^n$ for which the mean-square difference between the successive estimates $\hat{A}[n]$ and $\hat{A}[n-1]$ decays as $1/n^2$, the encoder/estimator structure (4.19) can be used to provide a host estimate $\check{A}[n]$ whose asymptotic MSE equals that of $\hat{A}[n]$. In particular, if the mean-square difference between successive sensor estimates

$\hat{A}[n]$ and $\hat{A}[n-1]$ decays as $1/n^2$, *i.e.*, if

$$\lim_{n\to\infty} n^2 E\left[\hat{\epsilon}^2[n]\right] = \gamma\,\sigma_v^2\,, \tag{4.27}$$

where $0 < \gamma < \infty$ and

$$\hat{\epsilon}[n] \stackrel{\Delta}{=} \hat{A}[n] - \hat{A}[n-1]\,, \tag{4.28}$$

then, as shown in App. C.3, the residual error between the sensor estimate $\hat{A}[n]$ and the associated host estimate $\check{A}[n-1]$ has the form (4.21), implying that the asymptotic MSE of $\check{A}[n]$ is the same as the one corresponding to $\hat{A}[n]$. The optimal value of λ in (4.19) can be found by minimizing the associate $\beta(\lambda)$ in (4.21); specifically, as we also show in the appendix we have

$$\beta(\lambda) \approx \frac{\pi\,(\gamma+\lambda^2)^2}{8\,\lambda^2}\,, \tag{4.29}$$

so that

$$\min_{\lambda} \beta(\lambda) \approx \pi\,\gamma/2,$$

which is achieved for $\lambda = \sqrt{\gamma}$.

Under a mild set of conditions on the sensor noise PDF, the ML estimator $\hat{A}_{\mathrm{ML}}[n]$ based on observation of $\mathbf{s}^n$, has the property that it is asymptotically efficient with respect to $\mathcal{B}(A;\,\mathbf{s}^n)$, asymptotically Gaussian distributed, and also satisfies (4.27) [12]. In these cases, when the sensor estimate computed is the ML estimate $\hat{A}_{\mathrm{ML}}[n]$ formed from $\mathbf{s}^n$, the block diagrams in Figs. 4-3 and 4-4 describe a general algorithmic method for obtaining an asymptotically efficient encoding.

Fig. 4-8 also depicts the MSE performance of the host estimator of the method (4.19) in the case that the sensor estimate $\hat{A}[n]$ is the ML estimate, which in this Laplacian scenario is the median of the n observations $s[1]$, $s[2]$, $\cdots$, $s[n]$ and is asymptotically efficient with respect to $\mathcal{B}(A;\,s)/n$. As the dashed curve in the figure reveals, the associated host estimate $\check{A}[n]$ based on the encodings is also asymptotically efficient.

4.3.4 Uniformly Distributed Noise

As we have already mentioned in Section 4.3.2, the estimator/detector structure described by (4.19) possesses remarkable robustness. As an illustration of this fact, in this section we consider estimation in IID uniformly distributed noise. In this case, the first-order PDF of $v[n]$ is given by

$$p_v(v) = \begin{cases} \frac{1}{2\sqrt{3}\sigma_v} & \text{if } |v| < \sqrt{3}\sigma_v \\ 0 & \text{otherwise} \end{cases} .$$

This noise process does not belong to the admissible class we have defined in Chapter 2. As is well known, a Cramér-Rao bound for this estimation problem does not exist, consistent with the fact that there exist estimators $\hat{A}[n]$ of A based on $\mathbf{s}^n$, whose MSE decays faster $1/n$. For instance, the MSE of the following estimator

$$\hat{A}[n] = \frac{\max\{s[1], s[2], \cdots, s[n]\} + \min\{s[1], s[2], \cdots, s[n]\}}{2} \tag{4.30}$$

decays as $1/n^2$:

$$E\left[\left(\hat{A}[n] - A\right)^2\right] = \frac{6\,\sigma_v^2}{(n+1)(n+2)} . \tag{4.31}$$

Even though the residual error between $\hat{A}[n]$ and $\check{A}[n]$ from the encoder/estimator pair (4.19) decays at best as fast as $1/n^2$ (see Theorem 2), by proper choice of λ the pair (4.19) we can effectively achieve the performance of $\hat{A}[n]$ in (4.30), as we demonstrate next.

The simulated MSE performance of the host estimate from (4.19) for $\lambda = 1$ and where the sensor estimate $\hat{A}[n]$ is given by (4.30) is depicted in Fig. 4-9. Note that since the variance of $\hat{\epsilon}[n]$ defined in (4.28) decays faster than $1/n^2$, γ from (4.27) equals zero. Since for any $\lambda > 0$ the asymptotic residual error scaling β is given by (4.29) for $\gamma = 0$, in this example we have $\beta = \pi/8$. Consequently, the asymptotic MSE of the host estimate $\check{A}[n]$ can be approximated as

$$\begin{aligned} E\left[\left(\check{A}[n] - A\right)^2\right] &\leq E\left[\left(\hat{A}[n+1] - \check{A}[n]\right)^2\right] + E\left[\left(\hat{A}[n+1] - A\right)^2\right] \\ &\lesssim (6 + \pi/8)\,\sigma_v^2/n^2 . \end{aligned} \tag{4.32}$$

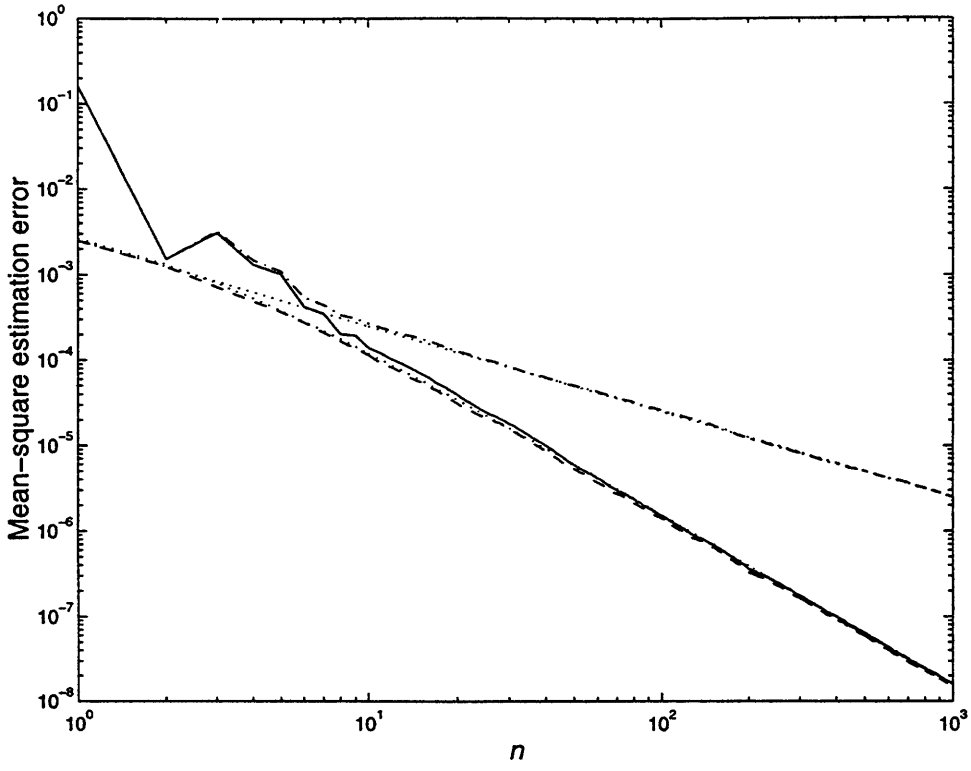


Figure 4-9: The dash-dot and solid curves show the host estimate MSE in uniformly distributed sensor noise, when the sample-mean and the estimator (4.30), respectively, are encoded at the sensor. For reference, the bound (4.32), σ_v^2/n, and the MSE of $\hat{A}[n]$ in (4.30) are depicted by the lower dotted, upper dotted, and dashed curve, respectively.

Combining (4.32) and (4.31) suggests that the encoder/decoder pair described by (4.19) and (4.30) incurs an asymptotic processing loss over the sensor estimate $\hat{A}[n]$ from (4.30) that is about

$$\mathcal{L}_{\text{Proc}}(A) \approx 1 + \frac{\pi}{48},$$

corresponding to only about 0.28 dB.

4.4 Network Extensions

Multi-sensor extensions of all the preceding single-sensor encoding and estimation algorithms can be constructed that retain the asymptotic optimality properties of the original single-sensor schemes. As in Chapter 3, we assume that $s_\ell[n]$, the nth observation collected

at the ℓth sensor, is given by

$$s_\ell[n] = A + v_\ell[n]\,,$$

where the sequences $v_\ell[n]$'s are independent IID noise sequences.

By designing the ℓth encoder according to the single-sensor principles and then properly combining the L symbol streams we can obtain asymptotically optimal estimates. As an illustration of the design of such multi-sensor extensions, we briefly consider fixed-rate encodings. Let $y_\ell[n]$ denote the sequence encoded at the ℓth sensor, $\check{A}_\ell[n]$ denote the asymptotically optimal host estimate resulting from using the encoding strategy (4.19) on the consistent sensor estimate $\hat{A}_\ell[n]$ formed at the ℓth sensor from

$$\mathbf{s}_\ell^n = \Big[s_\ell[1] \quad s_\ell[2] \quad \cdots \quad s_\ell[n]\Big]^T\,.$$

The ℓth encoder is depicted in Fig. 4-3, with $s[n]$, $y[n]$, $\hat{A}[n]$, and $\check{A}[n]$ replaced by $s_\ell[n]$, $y_\ell[n]$, $\hat{A}_\ell[n]$, and $\check{A}_\ell[n]$, respectively.

For simplicity and without loss of generality we consider the case where the original $\hat{A}_\ell[n]$'s are asymptotically efficient with respect to $\mathcal{B}\,(A;\ \mathbf{s}_\ell^n)$. In that case, the estimate

$$\check{A}[n] \triangleq \frac{1}{\sum_{\ell=1}^{L}\left[\mathcal{B}\,(0;\ s_\ell)\right]^{-1}} \sum_{\ell=1}^{L} \frac{\check{A}_\ell[n]}{\mathcal{B}\,(0;\ s_\ell)} \tag{4.33}$$

(where $\check{A}_\ell[n]$ is the estimate formed solely from the encodings of the ℓth sensor) provides an asymptotically efficient estimator of A from $\mathbf{s}_1^n$, $\mathbf{s}_2^n$, $\cdots$, $\mathbf{s}_L^n$. In the general case where the $\hat{A}_\ell[n]$'s are consistent but not necessarily efficient estimates with known MSE rates that are independent of the unknown parameter A, $\check{A}[n]$ from (4.33) provides an asymptotically optimal estimate provided we replace $\mathcal{B}\,(0;\ s_\ell)$ with $E\Big[(\hat{A}_\ell[n] - A)^2\Big]$.

Finally, in the special case that the PDFs of the sensor noises are identical *i.e.*, $p_{v_\ell}\,(x) = p_v\,(x)$ almost everywhere, the host estimator (4.33) reduces to

$$\check{A}[n] = \check{A}[n-1] + \frac{\lambda\,\sigma_v}{L\,n} \sum_{\ell=1}^{L} y_\ell[n]\,,$$

for $n \geq n_o$, and where we also used (4.19b). This decoder is also depicted in Fig. 4-4 provided we replace $y[n]$ with $\sum_{\ell=1}^{L} y_\ell[n]/L$.

Chapter 5

Encoding and Estimation with Quantizer Bias Control: Time-Varying Case

In Chapters 2 and 3 we have focused our attention on estimating a static signal from noisy measurements in the context of encoders composed of a control input added to each measurement prior to quantization. We have developed optimized encodings of the noisy measurements into digital sequences and asymptotically efficient estimators from these encodings for a number of scenarios of practical interest. Although our static case analysis has revealed a number of key characteristics of this signal estimation problem, the systems we have designed prove inadequate in cases where the information-bearing signal varies sufficiently fast to render the static signal assumption invalid across the observation interval used to form the estimates; in designing encoding strategies for the general time-varying case, we generally need to take into account the information-bearing signal characteristics, namely, the signal model and dynamics. However, as we show in this chapter, for a particular class of time-varying extensions, we can develop a rich class of encoding strategies and signal estimators by building on the principles that we have developed for the static case.

In this section we develop generalizations of the framework we have developed in Chapters 2 and 3 that encompass a number of time-varying information-bearing signals. Throughout this chapter we focus on information-bearing signals and sensor noises that are well modeled as Gaussian processes. In the general time-varying case, we usually have to rely

on coarse measurements from multiple sensors to obtain accurate signal estimates. This is clearly illustrated by considering the extreme case where the information-bearing signal is well modeled as an IID Gaussian process, and where the host is faced with the problem of estimating such a signal from encoded bits collected from a single sensor that measures this signal in statistically independent IID sensor noise. Since the information-bearing signal and the sensor noise are independent IID processes, for any fixed-rate binary encoding scheme (such as encoders of the form of quantizer bias control), at any given time instant n, the encoded bit can only provide information about the current signal sample. Furthermore, since past and future encodings do not provide any information for estimating the current signal sample, the problem of estimating any signal sample from the whole encoded sequence reduces to estimating the associated Gaussian random signal variable by observing a single encoded bit. Clearly, the ability of the host to estimate this Gaussian random variable based on a single bit is severely limited.

To overcome this problem, in this chapter we focus on signal estimation based on data collected from a network of sensors, each encoding one bit of information per measurement. In particular we focus on the special case where the information-bearing signal is perfectly correlated spatially over the sensor network, *i.e.*, at any time instant all sensors observe the same signal (in noise)[1]. In addition we assume that the sensor noise samples are independent in both time and space.

In Section 5.1 we present the class of time-varying signal models that we consider in this chapter. In Section 5.2 we state the figures of merit that we use to construct encoders and estimators for this class of time-varying signals. In Section 5.3 we present a number of methods that can be used to encode the noisy measurements at each sensor into bit streams. In Section 5.4 we sketch some of the methods that can be used to estimate the underlying information-bearing signal by intelligently fusing these bit streams at the host. Finally, In Section 5.5 we consider an example involving a simple signal model, which we use as a vehicle for illustrating the design and the performance characteristics of the schemes presented in Sections 5.3—5.4.

[1]We may also want to consider the dual problem, where the information-bearing signal is static in time, but partially correlated across the sensor network. Some of our analysis in this chapter carries through in this case, with appropriate modifications. Although beyond the scope of this thesis, a very interesting problem worth further investigation corresponds to the case where the signal samples are partially correlated in time and space. Indeed, a number of very interesting decentralized data fusion problems arise in that context; see [19, 10, 5, 6] and the references therein.

5.1 System Model

Throughout this chapter we focus our attention on the problem of estimating a single information-bearing signal $A[n]$ given by

$$A[n] = \mathbf{q}^T \mathbf{x}[n] \,, \tag{5.1a}$$

where

$$\mathbf{x}[n] = \begin{bmatrix} x[n] & x[n-1] & \cdots & x[n-R] \end{bmatrix}^T \tag{5.1b}$$

is a state-space vector that obeys the following dynamics

$$\mathbf{x}[n] = G\,\mathbf{x}[n-1] + \mathbf{h}\,u[n] \,, \tag{5.1c}$$

and where G is a known $R \times R$ matrix, $\mathbf{h}$ is a known $R \times 1$ vector, and $u[n]$ is a zero-mean IID Gaussian process of variance σ_u^2.

The linear state-space model (5.1) describing the dynamics of the information-bearing signal is fairly general and, as is well known, encompasses a number of broadly used signal models, including the autoregressive (AR), the moving-average (MA), and the autoregressive moving-average (ARMA) model [1]. For instance, an R-th order AR model of the form

$$A[n] = \sum_{i=1}^{R} a_i\, A[n-i] + u[n]$$

can be readily described via (5.1) by letting

$$\mathbf{q}^T = \begin{bmatrix} 1 & \mathbf{0}_{1\times R} \end{bmatrix}^T$$

$[\mathbf{h}]_1 = 1$, and

$$G = \begin{bmatrix} a_1 & a_2 & \cdots & a_R \\ I_{R-1} & & & \mathbf{0}_{R-1\times 1} \end{bmatrix} .$$

We consider an L-sensor scenario according to which the nth measurement at the ℓth

sensor is given by

$$s_\ell[n] = A[n] + v_\ell[n]\,, \tag{5.2}$$

where the sensor noise sequences $v_\ell[n]$ are statistically independent zero-mean IID Gaussian processes with variance σ_v^2, independent of the information-bearing signal $A[n]$. At time n, the ℓth sensor encodes the measurement $s_\ell[n]$ by means of quantizer bias control, *i.e.*,

$$y_\ell[n] = \mathrm{sgn}\left(s_\ell[n] + w_\ell[n]\right), \tag{5.3}$$

where $y_\ell[n]$ and $w_\ell[n]$ denote the encoded bit and the control input used at the ℓth sensor at time n, respectively. For compactness, we rewrite the above encoding equation (5.3) in matrix form as

$$\begin{bmatrix} y_1[n] \\ y_2[n] \\ \vdots \\ y_L[n] \end{bmatrix} = \begin{bmatrix} \mathrm{sgn}\left(s_1[n] + w_1[n]\right) \\ \mathrm{sgn}\left(s_2[n] + w_2[n]\right) \\ \vdots \\ \mathrm{sgn}\left(s_L[n] + w_L[n]\right) \end{bmatrix}. \tag{5.4}$$

Our objective is to design the control inputs $w_\ell[n]$ used at the sensors and the associated estimators at the host based on the state-space model given by (5.1), (5.2) and (5.4), so as to enable the host to obtain accurate signal estimates.

5.2 Performance Measures

Consistent with our previous developments, to address the quality of the encoding and the associated estimator we compare its performance against the one from the original measurements in the form of $s_\ell[n]$, *i.e.*,

$$\begin{bmatrix} s_1[n] \\ s_2[n] \\ \vdots \\ s_L[n] \end{bmatrix} = \mathbf{1}_{L\times 1}\,\mathbf{q}^T\,\mathbf{x}[n] + \begin{bmatrix} v_1[n] \\ v_2[n] \\ \vdots \\ v_L[n] \end{bmatrix}. \tag{5.5}$$

In particular, to design of the encoder at any given time instant n we use as performance metric the encoding (information) loss associated with estimation of $A[n]$ via

$$\mathbf{y}[n] \triangleq \begin{bmatrix} y_1[n] & y_2[n] & \cdots & y_L[n] \end{bmatrix}^T \tag{5.6}$$

instead of

$$\mathbf{s}[n] \triangleq \begin{bmatrix} s_1[n] & s_2[n] & \cdots & s_L[n] \end{bmatrix}^T . \tag{5.7}$$

Since we are interested in minimizing

$$\mathrm{MSE}(N) = \frac{1}{N} \sum_{n=1}^{N} E\left[\left(A[n] - \hat{A}[n]\right)^2\right] \tag{5.8}$$

to construct the encodings we use as our criterion average rather than worst-case (information loss) performance. We use as our figure of metric for designing the encodings at time n the average information loss in terms of estimating $A[n]$ based on observation of the $L \times 1$ vector $\mathbf{y}[n]$ instead of the vector $\mathbf{s}[n]$:

$$\mathcal{L}(A[n]; n) \triangleq \frac{\bar{\mathcal{B}}\left(A[n],\ \mathbf{y}[n]\right)}{\bar{\mathcal{B}}\left(A[n],\ \mathbf{s}[n]\right)} . \tag{5.9}$$

Since the sensor noise sequences are independent IID processes, minimizing the encoding loss (5.9) for each n, also minimizes the average encoding loss over *all* n. Similarly, to assess the performance of the estimator, we use as our figure of metric the average MSE loss, defined as the MSE performance (5.8) based on the sequence $\{\mathbf{y}[n]\}_{n=1}^{N}$ divided by the associated MSE performance via the sequence $\{\mathbf{s}[n]\}_{n=1}^{N}$.

As in the static case, both the design and performance of the systems employing quantizer bias control is dictated by the available freedom and the processing complexity in forming $w[n]$, as well as the number of sensors in the network. However, in this time-varying case the encoding performance of (5.4) also depends on the particular signal characteristics (5.1). As we show, however, in many cases of practical interest there exist naturally suited measures of SNR that can be used to describe the encoding and estimation performance. In particular, due to the perfect spatial correlation of the signal across the sensor sensor, to design the encoding at any time instant, it is convenient to view these encodings obtained

from all the sensors in the network as being equivalent to a temporal sequence of encodings of a static signal obtained from a single sensor. We first consider the design of the encoder and consequently address the estimation problem.

5.3 Encoding Algorithms

In this section we focus on designing encoding strategies based on quantizer bias control characterized by the control sequences $w_\ell[n]$. As we demonstrate, we can build on the principles we developed for the static case to develop a rich class of efficient encoding strategies for time-varying signals. We next develop encoders employing pseudo-noise control inputs, control inputs based on feedback and finally, combinations of pseudo-noise and feedback; similar strategies can be developed for control inputs known to the host, as well as any combinations thereof with pseudo-noise and feedback-based control inputs.

5.3.1 Pseudo-noise Control Inputs

In this section we consider the case where the control input sequences $w_1[n]$, $w_2[n]$, $\cdots$, $w_L[n]$ in (5.3) are statistically independent IID Gaussian processes, each with power level σ_w^2. The objective is to select the pseudo-noise power level σ_w^2 so as to minimize the average information loss of the form (5.9) that occurs when estimating $A[n]$ (for a fixed n) based on the $L \times 1$ vector $\mathbf{y}[n]$ in (5.6) instead of $\mathbf{s}[n]$ in (5.7), and where $y_\ell[n]$ denotes the output of the encoder (5.3) using binary quantizer bias control on $s_\ell[n]$ given by (5.2).

The dual interpretation of the signal encodings obtained at a given time instant from the sensor network as a temporally encoded sequence of a static signal obtained from a single sensor is extremely convenient, since it readily allows us to exploit the encoding principles we developed for the static case. As expected from the static case analysis, at any given time n, for pseudo-noise control inputs the optimal pseudo-noise level and the associate encoding performance are also functions of the signal power level, namely,

$$\sigma_A^2[n] \triangleq \text{var}\left(A[n]\right) ,$$

and the sensor noise level σ_v^2. In particular, following the analysis for pseudo-noise control inputs presented in Section 3.2.2, the average information loss (5.9) for a given signal strength $\sigma_A[n]$, sensor noise level σ_v, and pseudo-noise level σ_w can be denoted as

$\bar{\mathcal{L}}(\sigma_A[n], \sigma_v, \sigma_w)$ and is thus given by (3.13), where the average encoding performance is given by $\bar{\mathcal{B}}(\sigma_A[n], \sigma_v, \sigma_w)$, defined in Section 3.2.2.

Note that since the system (5.1) is LTI, $A[n]$ is a zero-mean Gaussian random variable whose variance converges to a constant σ_A^2 as $n \to \infty$. Thus, in steady state $\bar{\mathcal{L}}(\sigma_A[n], \sigma_v, \sigma_w)$ is independent of n. We are interested in the steady-state solution, *i.e.*, we wish to select σ_w so as to minimize the associated average information loss, *i.e.*,

$$\sigma_w^{\text{opt}} = \arg\min_{\sigma_w} \bar{\mathcal{L}}(\sigma_A, \sigma_v, \sigma_w) \ . \tag{5.10}$$

The optimal steady-state pseudo-noise level is then readily given by (3.22) where $\sigma^{\text{opt}}(1)$ is given by (3.20). The associated optimal average information loss is then given by $\bar{\mathcal{L}}^{\text{pn}}(\bar{\chi})$ from (3.23) where

$$\bar{\chi} \triangleq \frac{\sigma_A}{\sigma_v} \ .$$

Comparison of (3.23) and (3.24) reveals that proper use of pseudo-noise across the network improves the encoding efficiency over simply quantizing the measurements especially at high SNR $\bar{\chi}$; in particular, for large $\bar{\chi}$ the information loss (5.9) can be made to grow as slow as $\bar{\chi}^2$ by proper selection of the pseudo-noise power level.

5.3.2 Encodings Based on Feedback

Similarly, we may consider cases where feedback from the host to each sensor in the network is available, and takes the form

$$w_\ell[n] = w[n] = f(\mathbf{y}[k]; k < n) \ . \tag{5.11}$$

We would like to determine the performance limits of these strategies, and, in addition, to select $f(\cdot)$ so as to optimize the quality encodings.

To assess the performance limits of feedback methods it is convenient to consider the following alternative description of $A[n]$ in (5.1)

$$A[n] = \mathbf{q}^T G\, \mathbf{x}[n-1] + \mathbf{q}^T \mathbf{h}\, u[n] \ . \tag{5.12}$$

As we may recall from the static case analysis, the encoding performance is optimized if the encoder operates close to the quantizer threshold; ideally, we would like to use a control input $w[n]$ based on past encoded values that is as close to $-A[n]$ as possible. However, since feedback at time n can only depend on past observed encoded bits, *i.e.*, $\mathbf{y}[k]$ for $k < n$, we can only hope to accurately predict the component $\mathbf{q}^T G\,\mathbf{x}[n-1]$ of $A[n]$ in (5.12); the term $\mathbf{q}^T \mathbf{h}\, u[n]$ in (5.12) can not be predicted (and "subtracted" off) via feedback from past encodings, since this signal component is statistically independent of all past observation, *i.e.*, independent of all $\mathbf{s}[k]$ for $k < n$.

Assuming that the term $\mathbf{q}^T G\,\mathbf{x}[n-1]$ can be accurately predicted via feedback and subtracted from the measurement at each sensor, at any time instant n the information loss across the array is governed by the unpredictable component $\mathbf{q}^T \mathbf{h}\, u[n]$. Consequently, the information loss in the encodings via feedback is lower bounded by $\bar{\mathcal{L}}^{\text{free}}(\bar{\chi}_{\text{fb}})$ given by (3.24), where

$$\bar{\chi}_{\text{fb}} \triangleq \frac{\sigma'_u}{\sigma_v} \tag{5.13}$$

and where σ'^2_u is the power level of the term $\mathbf{q}^T \mathbf{h}\, u[n]$ of $A[n]$ in (5.12) (which cannot be predicted from the past)

$$\sigma'_u = \sigma_u\, \mathbf{q}^T \mathbf{h}\, \mathbf{h}^T \mathbf{q}\,. \tag{5.14}$$

The accuracy within which we can approach the bound $\bar{\mathcal{L}}^{\text{free}}(\bar{\chi}_{\text{fb}})$ depends on how accurately we can estimate the term $\mathbf{q}^T G\,\mathbf{x}[n-1]$ of $A[n]$ in (5.12) based on past observations. Since the estimate improves with the number of observations, we can get arbitrarily close to this bound provided we have enough *spatial* observations for each n, *i.e.*, provided that L is large enough.

For small enough feedback SNR $\bar{\chi}_{\text{fb}}$ in (5.13), encodings of the form (5.11) may incur only a small loss in performance. However, for high $\bar{\chi}_{\text{fb}}$ performance degrades very rapidly as (3.24) reveals. For this reason, we may consider joint use of feedback and pseudo-noise to improve performance for large $\bar{\chi}_{\text{fb}}$.

5.3.3 Joint Use of Pseudo-noise and Feedback

Joint use of pseudo-noise and feedback can provide significant performance improvements over using feedback or pseudo-noise alone. In this section we consider control inputs of the form

$$w_\ell[n] = w_{\text{fb}}[n] + \sigma_w \, \tilde{w}_\ell[n] \tag{5.15a}$$

where the sequences $\tilde{w}_\ell[n]$ are statistically independent IID zero-mean unit-variance Gaussian processes, and the feedback sequence $w_{\text{fb}}[n]$ is to be properly selected as a function of all past observations, *i.e.*,

$$w_{\text{fb}}[n] = f(\mathbf{y}[k]; k < n) \ . \tag{5.15b}$$

According to our development in the preceding section, we may use the feedback term (5.15b) to predict and "cancel" out the term $\mathbf{q}^T G \, \mathbf{x}[n-1]$ from $A[n]$, and then use pseudo-noise term to optimize the encoding in terms of the novel term. Provided that there are enough sensors to form accurate estimates of $\mathbf{q}^T G \, \mathbf{x}[n-1]$ from $\mathbf{y}[k]$ for $k < n$, to minimize the encoding loss we simply have to select the pseudo-noise level σ_w according to (3.22) for σ_A replaced by σ'_u. The resulting encoding strategy (5.15) can be used to achieve encodings whose information loss (5.9) grows as slow as quadratically with $\bar{\chi}_{\text{fb}}$ in (5.13); specifically, at high SNR $\bar{\chi}_{\text{fb}}$ the encoding performance is given by (3.19) for $\bar{\chi}$ replaced by $\bar{\chi}_{\text{fb}}$. As expected, joint use of pseudo-noise and feedback provides advantages over using only feedback or only pseudo-noise separately; since $\bar{\chi}_{\text{fb}} \leq \bar{\chi}$ and since $\bar{\mathcal{L}}^{\text{pn}}(\cdot)$ is an increasing function of its argument we have

$$\bar{\mathcal{L}}^{\text{pn}}(\bar{\chi}_{\text{fb}}) \leq \bar{\mathcal{L}}^{\text{pn}}(\bar{\chi}) \ , \tag{5.16}$$

revealing that the encoding performance of networks exploiting feedback and pseudo-noise is in general superior to that of networks exploiting pseudo-noise alone. In fact, since $\bar{\mathcal{L}}^{\text{pn}}(\cdot)$ is a strictly increasing function, equality in (5.16) is achieved if and only if $\bar{\chi}_{\text{fb}} = \bar{\chi}$, *i.e.*, if and only if $A[n]$ is an IID process. As (3.25) reveals we must also have

$$\bar{\mathcal{L}}^{\text{pn}}(\bar{\chi}_{\text{fb}}) \leq \bar{\mathcal{L}}^{\text{free}}(\bar{\chi}_{\text{fb}}) \ , \tag{5.17}$$

which suggests that joint use of feedback and pseudo-noise is advantageous over using feedback alone, if and only if $\sigma_w^{\text{opt}}(\sigma_u')$ in (3.22) equals zero, *i.e.*, if and only if $\bar{\chi}_{\text{fb}} < 1/\sigma^{\text{opt}}(1)$.

5.3.4 Other Encoding Strategies

We can similarly develop encoding strategies that are based on the use of distinct known control inputs across the network, and their combinations with pseudo-noise and/or feedback-based control inputs. For instance, we can design encodings using control inputs exploiting pseudo-noise, feedback, and known components which can provide improvements over control inputs based on joint use of pseudo-noise and feedback for large $\bar{\chi}_{\text{fb}}$. Specifically, by exploiting feedback we can effectively cancel out the term $\mathbf{q}^T G\,\mathbf{x}[n-1]$ in (5.12). By associating with the ℓth sensor a known predetermined quantizer bias $w_{\text{kn}}[n;\ell]$, and by using pseudo-noise inputs with smaller (optimized) σ_w, we can obtain performance improvements and make the average information loss to effectively grow as slow as linearly with $\bar{\chi}_{\text{fb}}$.

5.4 Signal Estimation

To illustrate some of the techniques that can be exploited to design effective estimators of time-varying signals based on encodings obtained via quantizer bias control, it is first convenient to consider estimation based on the original unconstrained observations $s_\ell[n]$ in (5.5). Let's consider estimation of $\hat{A}[k]$ based on observation of $\{\mathbf{s}[m]\}_{m\le n}$, and, in particular, let's focus on the case $k = n$. Due to the statistical independence of the IID noise components in $\mathbf{v}[n]$ and the form of (5.5), the sequence

$$\hat{s}[n] = \frac{1}{L}\sum_{\ell=1}^{L} s_\ell[n] \tag{5.18}$$

forms a sequence of sufficient statistics for estimating $A[n]$ based on $\mathbf{s}[n]$ [1]. Moreover, $\hat{s}[n]$ is the ML estimate of $A[n]$ based on $\mathbf{s}[n]$. Equivalently, we may replace the measurement equations (5.5) by the single measurement equation

$$\hat{s}[n] = A[n] + \hat{v}[n]\,, \tag{5.19}$$

where $\hat{v}[n]$ is a zero-mean IID Gaussian process with variance σ_v^2/L; since the sequence $\hat{s}[n]$ in (5.18) is a sequence of sufficient statistics for estimating $A[n]$ from $\mathbf{s}[n]$, optimized estimators for the model (5.1), (5.2) are the same as for the model (5.1), (5.19). We may exploit this observation to design signal estimators of $A[n]$ based on the encodings $\mathbf{y}[n]$; specifically, we can replace the L measurement equations (5.4) by a single measurement equation arising from making an estimate of the nth sample $A[n]$ based on the nth $L \times 1$ observation vector $\mathbf{y}[n]$.

In order to reduce the L sensor measurement equations (5.4) into a single equation we consider the use of the MAP estimate of $A[n]$ based on observation of $\mathbf{y}[n]$. In each particular encoding case our objective is to obtain a single "measurement" equation relating the MAP estimate of the sample $A[n]$ based on the $L \times 1$ vector sample $\mathbf{y}[n]$, and the signal $A[n] \sim \mathcal{N}(0, \sigma_A^2)$, and use that to design an appropriate Kalman Filter for signal estimation.

5.4.1 Pseudo-noise Control Inputs

For pseudo-noise control inputs the MAP estimator of $A[n]$ given $\mathbf{y}[n]$ is given via the EM algorithm (3.27) by replacing $\mathbf{y}^N$ with $\mathbf{y}[n]$ and N with L, where σ_A^2 is the steady-state variance of $A[n]$ (and where a few additional minor modifications are required). Specifically, the resulting algorithm takes the following form:

$$\hat{A}_{\mathrm{EM}}^{(k+1)}[n] = \frac{1}{1+\frac{\sigma_\alpha^2}{L\sigma_A^2}} \left[\hat{A}_{\mathrm{EM}}^{(k)}[n] + \frac{\sigma_\alpha \left[\mathcal{K}_{Y_1}(\mathbf{y}[n]) - L\,Q\left(z^{(k)}[n]\right)\right] \exp\left(-\left(z^{(k)}[n]\right)^2/2\right)}{\sqrt{2\pi}\, L\, Q\left(z^{(k)}[n]\right)\left[1 - Q\left(z^{(k)}[n]\right)\right]} \right], \tag{5.20a}$$

where

$$z^{(k)}[n] = \frac{\hat{A}_{\mathrm{EM}}^{(k)}[n]}{\sigma_\alpha}, \tag{5.20b}$$

$\sigma_\alpha = \sqrt{\sigma_v^2 + \sigma_w^2}$, and where $\hat{A}_{\mathrm{MAP}}[n]$ is given by

$$\hat{A}_{\mathrm{MAP}}[n] = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}[n]\,. \tag{5.20c}$$

For large enough L the MSE loss of this algorithm in terms of estimating $A[n]$ based on $\mathbf{y}[n]$ instead of $\mathbf{s}[n]$ effectively achieves the encoding information loss (5.9). Given that for

any given value of $A[n]$, the MAP estimate becomes asymptotically Gaussian with mean $A[n]$ and variance $\mathcal{L}(A[n])\,\sigma_v^2/L$, we may view $\hat{A}_{\rm MAP}[n]$ as

$$\hat{y}[n] = \hat{A}_{\rm MAP}[n] = A[n] + \hat{v}_y[n] \tag{5.21}$$

where the sequence $\hat{v}_y[n]$ is Gaussian with mean zero and variance

$$\sigma_{\hat{v}_y}^2[n] = \mathcal{L}(A[n])\,\sigma_v^2/L\ . \tag{5.22}$$

Note that the "equivalent" sensor noise variance $\sigma_{\hat{v}_y}^2[n]$ at time n is a function of the signal value $A[n]$ at time n. Assuming that the pseudo-noise power level has been optimally selected, we can approximate the variance of $\hat{v}_y[n]$ as

$$\sigma_{\hat{v}_y}^2 = \bar{\mathcal{L}}\,(\bar{\chi})\ \sigma_v^2/L\ , \tag{5.23}$$

where $\bar{\mathcal{L}}\,(\bar{\chi})$ is the average loss over all possible values of $A[n]$ for $\bar{\chi} = \sigma_A/\sigma_v$.

We can then design a Kalman filter[2] for the model (5.1), (5.21) and (5.23), namely [1],

$$\hat{\mathbf{x}}[n|n] = G\,\hat{\mathbf{x}}[n-1|n-1] + \boldsymbol{\mu}[n]\,(\hat{y}[n] - \mathbf{q}^T G\,\hat{\mathbf{x}}[n-1|n-1]) \tag{5.24a}$$

$$\boldsymbol{\mu}[n] = \Lambda_x[n|n-1]\,\mathbf{q}\,(\mathbf{q}^T\Lambda_x[n|n-1]\,\mathbf{q} + \sigma_{\hat{v}_y}^2\,I)^{-1} \tag{5.24b}$$

$$\Lambda_x[n|n] = (I - \boldsymbol{\mu}[n]\,\mathbf{q}^T)\,\Lambda_x[n|n-1] \tag{5.24c}$$

$$\Lambda_x[n|n-1] = G\,\Lambda_x[n-1|n-1]\,G^T + \sigma_u^2\,\mathbf{h}\,\mathbf{h}^T \tag{5.24d}$$

initialized with $\hat{\mathbf{x}}[-1|-1] = \mathbf{0}$ and $\Lambda_x[-1|-1] = \sigma_u^2\,\mathbf{I}$.

It is worthwhile to note that, in general, $\mathbf{q}^T\Lambda_x[n|k]\mathbf{q}^T$ provides only an estimate of $\lambda_A[n|k]$, the MSE of the estimate of $A[n]$ given all observations up to and including time k, since the model used to construct the Kalman Filter (5.24) is only an approximate one; *cf.*, Eqns. (5.22) and (5.23).

[2]In fact, we can also design an Extended Kalman Filter for the original nonlinear state-space model given by (5.1), (5.21) and (5.22).

5.4.2 Estimation via Feedback

We can use a similar approach to design control input strategies based on feedback and associated signal estimators. Specifically, we can use the MAP estimate of $A[n]$ based on $\mathbf{y}[n]$ for any control input $w_{\text{fb}}[n]$. To cancel out the term $\mathbf{q}^T G\,\mathbf{x}[n-1]$, we can select this feedback term as

$$w_{\text{fb}}[n] = -\mathbf{q}^T G\,\hat{\mathbf{x}}[n-1|n-1]\;, \tag{5.25}$$

where $\hat{\mathbf{x}}[n-1|n-1]$ is our estimate of $\mathbf{x}[n-1]$ based on all past observations, *i.e.*, all available observations up to and including time $n-1$. Similarly to (5.21) we may view as our measurement equation

$$\hat{y}[n] = \hat{A}_{\text{MAP}}[n] = A[n] + \hat{v}_y[n], \tag{5.26}$$

where $\hat{A}_{\text{MAP}}[n]$ is given by (5.20a) and (5.20c) with

$$z^{(k)}[n] = \frac{\hat{A}^{(k)}_{\text{EM}}[n] + w_{\text{fb}}[n]}{\sigma_\alpha}\,, \tag{5.27}$$

and $\sigma_\alpha = \sigma_v$. Again we approximate the zero-mean strictly white nonstationary noise source $\hat{v}_y[n]$ with a zero-mean IID process of power level given by

$$\sigma^2_{\hat{v}_y} = \bar{\mathcal{L}}^{\text{free}}\left(\bar{\chi}_{\text{fb}}\right)\,\sigma_v^2/L\;, \tag{5.28}$$

We can then design a Kalman filter for the system model (5.1), (5.26) and (5.28); it is given by (5.24) where $\hat{y}[n]$ and $\sigma_{\hat{v}_y}$ are instead given by (5.26) and (5.28), respectively.

5.4.3 Estimation in Presence of Feedback and Pseudo-noise

We can easily extend the estimators of the previous sections to take into account use of both pseudo-noise and feedback. Specifically, as suggested in Section 5.3.3, we may use feedback of the form (5.25). We may use as our measurement equation (5.26), where $\hat{A}_{\text{MAP}}[n]$ is the MAP estimate of $A[n]$ based on observation of $\mathbf{y}[n]$ and is given by (3.27) with minor modifications. Specifically, it is given by where (5.20a), (5.20c), and (5.27) with

$\sigma_\alpha = \sqrt{\sigma_v^2 + \sigma_w^2}$. In that case, the power level of the "noise" process $\hat{u}_y[n]$ is given by

$$\sigma_{\hat{v}_y}^2 = \mathcal{L}(A[n],\, \sigma_\alpha)\, \sigma_v^2/L \;, \tag{5.29}$$

which in the case that the pseudo-noise power level is optimally selected is given by

$$\sigma_{\hat{v}_y}^2 = \bar{\mathcal{L}}^{\rm pn}\,(\bar{\chi}_{\rm fb})\; \sigma_v^2/L \;, \tag{5.30}$$

for large $\bar{\chi}_{\rm fb}$. Especially in the case that the pseudo-noise power level is optimally selected the measurement model (5.26) where $\hat{v}_y[n]$ is assumed an IID process of variance given by (5.29) is a reasonably accurate model for the original measurements equations. The Kalman filtering solution for this approximate model is given by (5.24), where $\sigma_{\hat{v}_y}$ and $\hat{y}[n]$ are given by (5.29) and (5.26), respectively.

5.5 Encoding and Estimation of an AR(1) process

As a brief illustration of the construction of the encoding strategies, the associated estimators, and their performance characteristics, we next consider a simple example involving estimation of a first order AR process given by

$$A[n] = \rho A[n-1] + \sqrt{1-\rho^2}\,\sigma_A \tilde{u}[n] \tag{5.31}$$

where $\tilde{u}[n]$ is a zero-mean unit-variance IID Gaussian process, and $0 \leq \rho \leq 1$. As is well known, for the parametric model (5.31), the parameter ρ can be viewed as a rough measure of signal bandwidth; for $\rho = 1$, $A[n]$ in (5.31) reduces to the static case which we have considered in detail in earlier chapters; for $\rho = 0$, $A[n]$ in (5.31) is a zero-mean IID Gaussian process with power level σ_w^2. Fig. 5-1 shows a typical sample path for an intermediate value of ρ.

Let's consider a scenario involving a distributed network of L sensors measuring $A[n]$ in statistically independent IID sensor noises as in (5.2) and employing binary quantizer bias control. As suggested in Section 5.3.3, joint use of feedback and pseudo-noise is in general superior over using feedback alone. This is clearly illustrated in Fig. 5-2, where we consider encodings of the form (5.15) for various σ_w levels, for a network of $L = 10^3$ sensors. As the

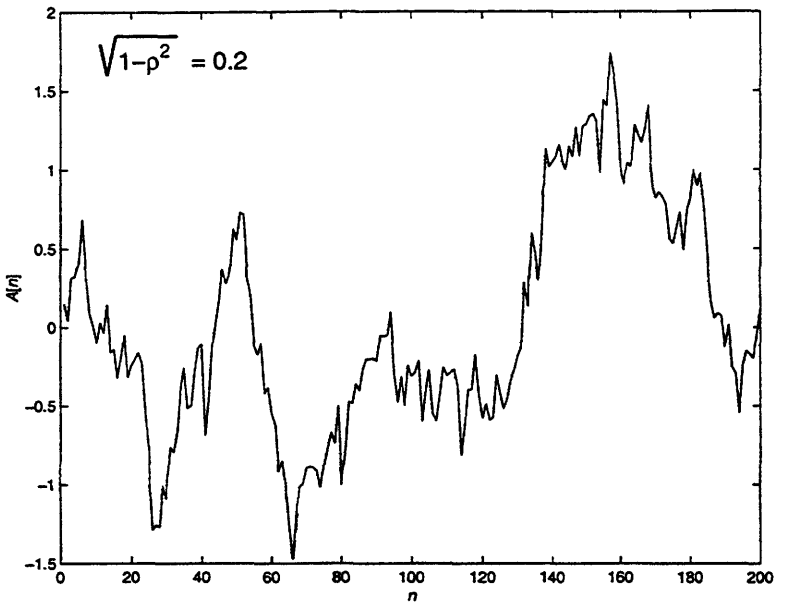


Figure 5-1: Sample path of an AR(1) process with dynamics given by (5.31), where $\sqrt{1-\rho^2} = 0.2$, $\sigma_A = 1$.

figure reveals, there is an optimal power level in terms of minimizing the associated MSE loss. The optimal power level is in fact very accurately predicted by (3.22) for σ_A replaced by σ'_u from (5.14).

Fig. 5-3 depicts the performance of this encoding strategy as a function of the "bandwidth" parameter ρ. As the figure reveals, in the static case ($\sqrt{1-\rho^2} = 0$) feedback alone provides the optimal encoding loss ($\approx$ 2 dB). At the other extreme, *i.e.*, $\sqrt{1-\rho^2} = 0$, feedback does not provide any encoding benefits; each signal sample $A[n]$ is independent of all past and future signal samples so we can not rely of past encodings to effectively predict any future $A[n]$ samples. On the other hand, suitable use of pseudo-noise across the network can provide performance benefits. And for intermediate ρ values, joint use of feedback and pseudo-noise provides performance improvements over using feedback or pseudo-noise alone.

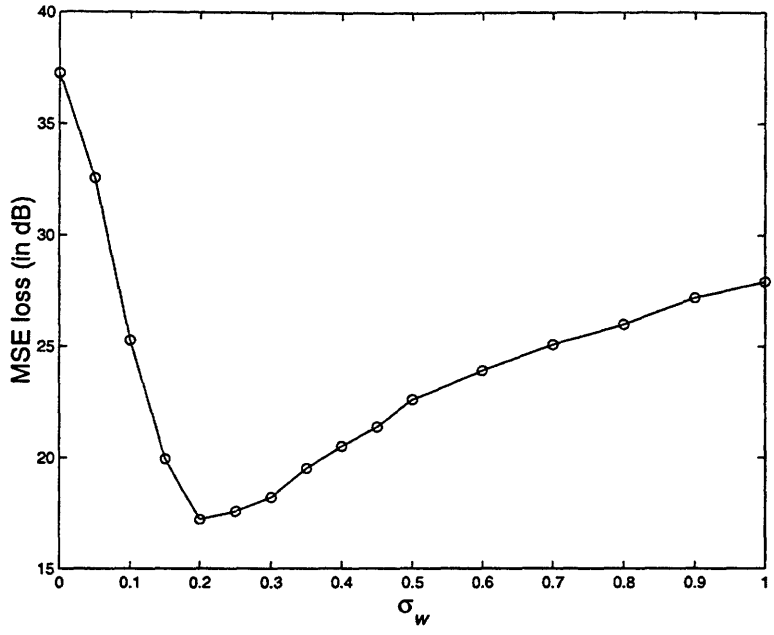


Figure 5-2: MSE loss in estimating an AR(1) process with dynamics given by (5.31), where $\sqrt{1-\rho^2} = 0.2$, $\sigma_A = 1$, based on a network of sensors using quantizer bias control according to (5.15), and where $\sigma_v = 0.1$.

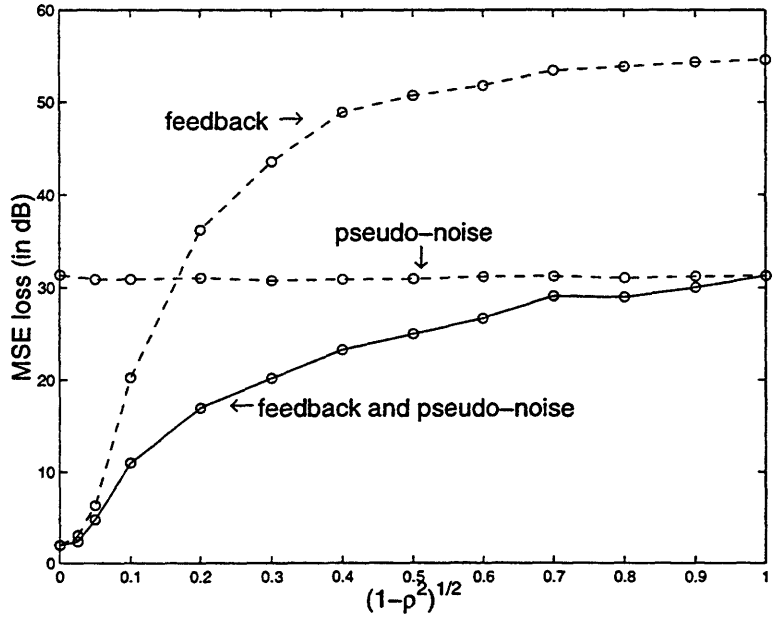


Figure 5-3: MSE loss in estimating an AR(1) process with dynamics given by (5.31), as a function of $\sqrt{1-\rho^2} = 0.2$ for $\sigma_A = 1$, based on a network of sensors using quantizer bias control, for pseudo-noise (dashed), feedback-based (dash-dot), and jointly optimized pseudo-noise and feedback-based control inputs (solid). Sensor noise level: $\sigma_v = 0.1$.

Chapter 6

Contributions and Future Directions

In this thesis we have focused on signal estimation from noisy measurements, where system constraints force us to rely on a quantized description of the noisy measurements. We have developed a framework for designing encodings of the noisy measurements into efficient digitized descriptions and optimized signal estimators from these encodings for a number of important scenarios with various encoder complexity characteristics.

As a main contribution of this thesis, we have introduced encodings of the form of what we refer to as quantizer bias control. For the static signal case, we have developed optimized encodings for a variety of important scenarios that may arise in practice, together with associated estimators which are asymptotically achieve the optimal performance from these encodings. Specifically, we have developed a framework for evaluating these quantizer-based systems by means of a figure of merit which we refer to as the information loss; it is defined as the increase in dB that is incurred in the Cramér-Rao bound for unbiased estimates by a particular type of additive control input and a given M-level quantizer. In general, for control-free systems the performance rapidly degrades with peak signal-to-noise ratio (SNR) χ, which is defined as the ratio of the parameter dynamic range to the sensor noise power level. In particular, as we have shown, for a wide class of IID sensor noises the worst-case information loss grows faster than χ^2 if no control input is used.

We have considered a number of important scenarios that may arise in practice which differ in terms of the available knowledge about the control waveform for estimation and

the associated freedom in the control input selection. If only the statistical characterization of the control input can be exploited for estimation, we have shown that pseudo-noise control inputs can provide significant performance benefits, in the sense that the worst-case information loss can be made to grow as slow as quadratically with SNR. If knowledge of the particular control input is exploited for estimation, even higher performance can be achieved. In particular, we have developed methods for selecting the control input from a suitably designed class of periodic waveforms, for which the worst-case information loss grows linearly with SNR. Finally, for cases where feedback is available we have developed control waveform selection strategies and corresponding computationally efficient estimators that asymptotically achieve the best possible performance for quantizer-based systems with additive control inputs. Specifically, these estimators achieve the minimum possible information loss for the associated quantizer-based system which is independent of SNR. It is worth emphasizing that these performance characteristics are exhibited by any M-level quantizer and a wide class of IID sensor noises. Furthermore, our methodology easily generalizes to scenarios involving networks of sensors employing quantizer bias control.

For all encoder complexity scenarios we considered, we have shown that optimized encodings have the same asymptotic characteristics even when the figure of merit is average (rather than worst-case) performance, *i.e.*, when there is prior information regarding the relative likelihood of the signal values. Furthermore, these asymptotic performance rates remain unaffected even if the sensor noise power level in the original measurements is unknown.

Although quantizer bias control encoders exploiting feedback can be constructed whose performance does not degrade with SNR, in general these systems incur a small information loss. This loss in performance is an inherent limitation of all encoders employing quantizer bias control and can only be eliminated by allowing more freedom in the encoder design. For cases where such freedom is available, we have developed a framework for designing efficient refinable encoding descriptions and estimators from these descriptions which asymptotically achieve the performance of any estimator that could be computed at the encoder from the original noisy measurements. In the event that the estimate computed at the encoder is asymptotically efficient with respect to the original sensor measurements, these encoder and estimator pairs have the attractive property that they achieve asymptotically optimal performance, *i.e.*, the resulting estimate based on the encodings asymptotically achieves

the best possible performance based on the original sensor measurements.

A very important extension of the encoding and estimation strategies involves developing efficient encodings of noisy measurements of time-varying information-bearing signals obtained at multiple sensors. Although the framework we have developed for the static case is in general inadequate for efficient encoding in the time-varying case, we have shown that we can exploit the key encoding principles used in the static analysis to develop a rich class of encoding strategies for time-varying signals. In particular, pseudo-noise, deterministic, and feedback-based control inputs can be effectively combined to provide improved performance over encodings strategies relying on only one of these types of control inputs. We have shown that in all cases performance is intricately linked to an appropriate measure of signal-to-noise ratio which depends on the particular signal characteristics and the allowed freedom in the encoder design. In the same context, we have developed estimators that make use of static case estimation principles to transform the multi-sensor measurements into an equivalent sufficient "single measurement" characterization which enables the use of a Kalman filter based approach to estimation.

Although we have sketched a number of optimized strategies that can be used to encode and estimate noisy time-varying signals, there are a number of important issues that must be successfully addressed to make such schemes practical in the context of distributed sensor networks. Typical issues that may arise in practical wireless sensor networks include inherent delays in all encoding strategies that exploit feedback, as well as the signal variability that is often exhibited across a network of this form.

6.1 Future Directions

While a large class of problems have been addressed in this thesis, there are a number of very important extensions that warrant further investigation. Indeed, some of these problems have been identified within the appropriate thesis chapters. However, there are a number of other important future directions that are immediately suggested by this work as well as potential connections with other important problems arising in various other areas of research.

In the context of parameter estimation based on encodings via quantizer bias control, for instance, it is important to study the performance that is achievable based on finite-

length observation windows. Such analysis may be beneficial for a number of applications involving signal quantizers. In addition, in most of our analysis we have assumed that the sensor noises are IID processes. However, in many applications sensor noise samples are temporally or even spatially correlated.

An interesting future direction pertains to extending the optimized higher-complexity encoding schemes we have developed in Chapter 4 in the context of time-varying signals. An intriguing question pertains to determining the best possible performance that is achievable by any such system, as well as the magnitude of the performance losses introduced by constraining the encoding strategy to the simpler quantizer bias control methods we have developed.

The framework we have introduced may potentially provide insight in many other research areas. Indeed, this is one of the potentially most fascinating directions for future work. For instance, the framework we have developed appears naturally suited for evaluation A/D conversion of noisy analog signals. In this case, the A/D converter has the dual function of removing noise from the noisy analog signal and of constructing an accurate digitized estimate of the analog signal. Indeed, some of the systems we have developed in this thesis may be useful in designing high bit-rate A/D converter arrays. However, the practical constraints that dictate the design of these systems may differ, in general, from the ones that we have considered in depth in this thesis [20, 21, 29].

Dithered quantizers find use in a number of other applications such as reconstruction of bandlimited signals via coarse oversampling [36], and halftoning techniques for images [25, 27]. The objective in halftoning is to add pseudorandom patterns to an image signal before coarse quantization as a method of removing visual artifacts that occur from coarse quantization of image areas that exhibit small signal variation. A number of halftoning techniques, *e.g.*, [27], can be viewed as parallels of pseudo-noise quantizer-bias control encodings of the original image into coarsely quantized pixels where there is the additional constraint that the "estimator" to be used is our visual system. Further connections between halftoning techniques and the systems we have developed in this thesis in the context of constrained signal estimation have yet to be explored.

Perhaps, the most exciting and fruitful future directions of this thesis pertain to finding connections and forming ties between this work and other important problems that arise in other disciplines in science and engineering.

Appendix A

A.1 Worst-Case Information Loss for Control-free Signal Quantizers

In this appendix we show that the worst-case information loss of any signal quantizer grows faster than χ^2 for large χ in the absence of a control input. We first consider the case $M = 2$ and show by contradiction that $\mathcal{L}_{\max}^{\text{free}}(\chi) \neq o\left(\chi^2\right)$ as $\chi \to \infty$, *i.e.*, we show that

$$\lim_{\chi \to \infty} \frac{\mathcal{L}_{\max}^{\text{free}}(\chi)}{\chi^2} = 0 \tag{A.1}$$

cannot be true. Letting $\chi \to \infty$ is equivalent to fixing Δ and letting $\sigma_v \to 0^+$, since the control-free information loss for $M = 2$ is completely characterized by χ. Let $\mathcal{B}_{\max}(\Delta, \sigma_v; y)$ denote the worst-case Cramér-Rao bound for estimating A from one sample of the IID sequence $y[n]$, for $|A| < \Delta$, and noise level σ_v. Then, (A.1) implies that

$$\lim_{\sigma_v \to 0^+} \mathcal{B}_{\max}(\Delta, \sigma_v; y) = 0\,, \tag{A.2}$$

where we used (2.4), (2.7) and (2.8). However, (A.2) suggests that, as $\sigma_v \to 0^+$, we can estimate *any* A in $(-\Delta, \Delta)$ with infinite accuracy from an one-bit observation $y[n]$, which is not possible. Thus, (A.1) is false, *i.e.*, $\mathcal{L}_{\max}^{\text{free}}(\chi)$ has to grow at least as fast as χ^2.

Similarly, we can also show that $\mathcal{L}_{\max}^{\text{free}}(\chi)$ grows faster that χ^2, in the sense that $\mathcal{L}_{\max}^{\text{free}}(\chi) \neq O(\chi^2)$. We show this by first assuming that $\mathcal{L}_{\max}^{\text{free}}(\chi) = O(\chi^2)$, *i.e.*, that we can find $D < \infty$ and χ_o, such that for $\chi > \chi_o$ we have $\mathcal{L}_{\max}^{\text{free}}(\chi) \leq D\,\chi^2$, and arriving to a contradiction. The condition $\mathcal{L}_{\max}^{\text{free}}(\chi) = O(\chi^2)$ is equivalent to the statement that there

exists $D < \infty$ such that

$$\limsup_{\chi \to \infty} \frac{\mathcal{L}_{\max}^{\text{free}}(\chi)}{\chi^2} = D \, . \tag{A.3}$$

Again using (2.4) and (2.7)–(2.8) in (A.3) we obtain the following equivalent statement,

$$\limsup_{\sigma_v \to 0^+} \mathcal{B}_{\max}(\Delta, \sigma_v; y) = D' , \tag{A.4}$$

where $D' < \infty$. Since the sequence $y[n]$ is IID, (A.4) implies that as $\sigma_v \to 0^+$, the Cramér-Rao bound $\mathcal{B}\left(A;\ \mathbf{y}^N\right)$ is upper-bounded by D'/N, which goes to 0 as $N \to \infty$. However, for any $A \neq 0$, in the limit $\sigma_v \to 0^+$ we have $y[n] = \text{sgn}(A)$ with probability 1 for all n, which in turn implies that $\mathcal{B}\left(A;\ \mathbf{y}^N\right)$ cannot go to 0 as $N \to \infty$. Thus, we must have $D' = \infty$ in (A.4), which proves that the control-free worst-case information loss is not $O(\chi^2)$.

We can show that $\mathcal{L}_{\max}^{\text{free}}(\chi) \neq O\left(\chi^2\right)$ for signal quantizers with $M > 2$, by using our results for $M = 2$. Specifically, if Δ is fixed, in which case $\chi \to \infty$ is equivalent to $\sigma_v \to 0^+$, the arguments used for the $M = 2$ case still apply with minor modifications. Next consider fixing σ_v, in which case $\chi \to \infty$ is equivalent to $\Delta \to \infty$. As usual, let X_1, X_2, X_{M-1} denote the quantizer thresholds. By rescaling by $1/\Delta$, this problem can be mapped to an equivalent one where $\Delta' = 1$, $\sigma_v' = \sigma_v/\Delta \to 0^+$, and where the new quantizer thresholds are $X_1/\Delta, X_2/\Delta, X_{M-1}/\Delta$. The arguments used to show that $\mathcal{L}_{\max}^{\text{free}}(\chi) \neq O\left(\chi^2\right)$ in the $M = 2$ case still apply in this case with minor modifications.

A.2 Worst-Case Information Loss for Known Control Inputs

We first show that for any known control input scenario, the worst-case information loss grows at least as fast as χ. This is true for any sensor noise distribution and for any $M \geq 2$. For convenience, we denote by $p_w(\cdot)$ the empirical probability density function of the known sequence $w[n]$ [26]. The associated Cramér-Rao bound for estimating A based on $y[n]$ for a particular $p_w(\cdot)$ is given by

$$\mathcal{B}\left(A;\ \mathbf{y}^N,\ p_w(\cdot)\right) = \frac{1}{N}\left(E\left[\left\{\mathcal{B}\left(A + w;\ y\right)\right\}^{-1}\right]\right)^{-1} \tag{A.5}$$

where the expectation is with respect to $p_w(\cdot)$. For instance, if the periodic sequence (2.26) is represented by an empirical PDF consisting of K Kronecker delta functions located at $w[n]$ for $n = 0, 1 \cdots, K-1$ and each with area $1/K$, then (A.5) and (2.24) are equivalent.

For convenience, we consider the inverse of the Cramér-Rao bound in (A.5), namely, the Fisher information of A given $y[n]$. We denote the Fisher information in the control-free case by $\mathcal{F}(A;\ y)$. The worst-case Fisher information $\mathcal{F}_{\min}(\Delta;\ p_w(\cdot))$ for an input with an empirical PDF $p_w(\cdot)$ is defined as

$$\mathcal{F}_{\min}(\Delta;\ p_w(\cdot)) \triangleq \inf_{|A|<\Delta} E[\mathcal{F}(A+w;\ y)]$$

where the expectation is with respect to $p_w(\cdot)$. Consider the optimal selection of $p_w(\cdot)$, which results in maximizing $\mathcal{F}_{\min}(\Delta;\ p_w(\cdot))$, *i.e.*,

$$\mathcal{F}_{\text{opt}}(\Delta) \triangleq \max_{p_w(\cdot)} \mathcal{F}_{\min}(\Delta;\ p_w(\cdot)) \ .$$

The growth of the optimal worst-case information loss equals the growth of the inverse of the optimal worst-case Fisher information defined above.

We will make use of the fact that the control-free worst-case information loss grows strictly faster than χ^ρ for $\rho < 2$ (*cf.* the generalization of (A.1) to quantizers with $M \geq 2$). Without loss of generality we may set $\sigma_v = 1$, in which case $\Delta = \chi$. Since $\mathcal{B}(A;\ s)$ is independent of A (and thus Δ), the control-free worst-case Fisher information of A based on $y[n]$ decays faster than $1/\Delta^\rho$ for any $\rho < 2$, as Δ increases. Thus, there exist $D > 0$ and $\delta > 0$, such that for any $|A| > \delta$

$$\mathcal{F}(A;\ y) = [\mathcal{B}(A;\ y)]^{-1} < \min\left\{ D\,|A|^{-\rho},\ [\mathcal{B}(A;\ s)]^{-1} \right\} , \tag{A.6}$$

for any given $\rho < 2$. For convenience, we pick ρ so that $1 < \rho < 2$. Also, let

$$\mathcal{P}_k(A;\ p_w(\cdot)) \triangleq \int_{k\,\delta \leq |w+A| < (k+1)\,\delta} p_w(w)\ dw \ .$$

For any empirical PDF $p_w(\cdot)$ and any Δ satisfying $\Delta > \delta$, we must have

$$\inf_{|A|<\Delta} \mathcal{P}_k(A;\ p_w(\cdot)) < \frac{2\,\delta}{\Delta} \ . \tag{A.7}$$

We can establish (A.7) via proof by contradiction; if the inequality in (A.7) is reversed, for any A in $(-\Delta, \Delta)$ we have

$$\mathcal{P}_k\left(A;\, p_w\left(\cdot\right)\right) \geq \frac{2\,\delta}{\Delta}. \tag{A.8}$$

Let $A_j = j\,\delta$ for $j = 0, \pm 1, \cdots, \pm j_o$, where j_o is the largest index j satisfying $A_j < \Delta$. Note that $j_o > \Delta/(2\,\delta)$. Applying (A.8) for $A = A_j$, and summing over all j yields

$$\sum_{j=-j_o}^{j_o} \mathcal{P}_k\left(A_j;\, p_w\left(\cdot\right)\right) \geq \left(2\,j_o + 1\right)\frac{2\,\delta}{\Delta}, \tag{A.9}$$

which is a contradiction since the left hand size of (A.9) is upper-bounded by $2\int_w p_w\left(w\right)\,dw$, while $\left(2\,j_o + 1\right)\left(2\,\delta\right)/\Delta > 2$. We can similarly derive the following generalization of (A.7)

$$\inf_{|A|<\Delta} \sum_k \beta_k\, \mathcal{P}_k\left(A;\, p_w\left(\cdot\right)\right) < \frac{2\,\delta}{\Delta} \sum_k \beta_k\ , \tag{A.10}$$

where $\beta_k \geq 0$ and at least one of the β_k's is non-zero. We have

$$\begin{aligned}
\mathcal{F}_{\text{opt}}\left(\Delta\right) &= \max_{p_w(\cdot)} \inf_{|A|<\Delta} \sum_{k=0}^{\infty} \int_{k\,\delta \leq |w+A| < (k+1)\,\delta} p_w\left(w\right)\, \mathcal{F}\left(A + w;\, y\right)\, dw \\
&< \max_{p_w(\cdot)} \inf_{|A|<\Delta} \left(\left[\mathcal{B}\left(0;\, s\right)\right]^{-1} \mathcal{P}_0\left(A;\, p_w\left(\cdot\right)\right) + \sum_{k=1}^{\infty} \frac{D}{\left(k\,\delta\right)^{\rho}}\, \mathcal{P}_k\left(A;\, p_w\left(\cdot\right)\right) \right) && \text{(A.11a)} \\
&< \frac{2\delta}{\Delta}\left(\left[\mathcal{B}\left(0;\, s\right)\right]^{-1} + \frac{D}{\delta^{\rho}} \sum_{k=1}^{\infty} \frac{1}{k^{\rho}} \right) && \text{(A.11b)} \\
&\leq \frac{C\,\delta}{\Delta} && \text{(A.11c)}
\end{aligned}$$

where $C < \infty$, since $\sum_{k=1}^{\infty} k^{-\rho}$ is a convergent series for $\rho > 1$. To obtain (A.11a) and (A.11b) we used (A.6) and (A.10), respectively. As (A.11c) reveals, for large Δ the optimal worst-case information loss grows at least as fast as χ (since $\chi = \Delta$ for $\sigma_v = 1$).

We next show that simple periodic control input schemes can be constructed for which the worst-case information loss (for $N \to \infty$) grows linearly with χ. It suffices to consider signal quantizers with $M = 2$, since signal quantizer with $M > 2$ provide additional information and would thus perform at least as well. In particular, we next show that K-periodic waveforms given by (2.26), where K is given by (2.27) for a fixed $\lambda > 0$, achieve the opti-

mal growth rate for any admissible sensor noise and a symmetric two-level quantizer. Let $\bar{\mathcal{B}}(A;\,\sigma_v)$ denote the Cramér-Rao bound (2.14) with α replaced by v. Note that since

$$\bar{\mathcal{B}}(A;\,\sigma_v) = \sigma_v^2\,\bar{\mathcal{B}}(A/\sigma_v;\,1)\,, \tag{A.12}$$

we also have $\mathcal{B}_{\max}(\Delta;\,\sigma_v) = \sigma_v^2\,\mathcal{B}_{\max}(\Delta/\sigma_v;\,1)$, which in conjunction with (2.8) reveals that the associated information loss is completely characterized by the ratio $\chi = \Delta/\sigma_v$. Since K also solely depends on χ, we may fix $\Delta = 1$ without loss of generality. Note that the class (2.26) remains invariant to changes in σ_v. Hence, we may use $w[n;K]$ to denote the unique K-periodic sequence from the class (2.26) corresponding to $\Delta = 1$. For $\sigma_v < \lambda$, we have $3\,\lambda/\sigma_v > K$, and

$$\begin{aligned}
\mathcal{B}_{\max}(1,\sigma_v) &= \sup_{A\in(-1,1)} \frac{K}{\sum_{n=1}^{K}\left[\bar{\mathcal{B}}(A+w[n;K];\,\sigma_v)\right]^{-1}} && \text{(A.13a)}\\
&< \frac{3\lambda}{\sigma_v} \sup_{A\in(-1,1)} \min_{n\in\{1,2,\cdots,K\}} \bar{\mathcal{B}}(A+w[n;K];\,\sigma_v) && \text{(A.13b)}\\
&\le 3\,\lambda\,\sigma_v \sup_{A'\in(-1/\sigma_v,\,1/\sigma_v)} \min_{n\in\{1,2,\cdots,K\}} \bar{\mathcal{B}}(A'+w'[n;K];\,1) && \text{(A.13c)}\\
&\le 3\,\lambda\,\sigma_v \sup_{A\in(-1/\lambda,\,1/\lambda)} \bar{\mathcal{B}}(A;\,1)\,, && \text{(A.13d)}
\end{aligned}$$

where $w'[n;K] = w[n;K]/\sigma_v$, and where we used (A.12) to obtain (A.13c) from (A.13b). To verify (A.13d) from (A.13c), note that for any fixed A' in $(-1/\sigma_v,\,1/\sigma_v)$, the minimum of $\bar{\mathcal{B}}(A'+w'[n;K];\,1)$ over n is upper-bounded by $\bar{\mathcal{B}}(A'+w'[n';K];\,1)$, where n' is the value of n for which $|A'+w'[n;K]|$ is the smallest. Since the spacing $\delta_{w'}$ of the sawtooth waveform $w'[n;K]$ satisfies $\delta_{w'} = \delta_w/\sigma_v \le 2/\lambda$, $|A'+w'[n';K]|$ is upper-bounded by $\delta_{w'}/2 \le 1/\lambda$ for any $|A'| < 1/\sigma_v$, verifying (A.13d). Since $\mathcal{B}(A;\,s) \sim \sigma_v^2$ from (2.8) and by using (A.13d), the worst-case information loss for known $w[n]$ given by (2.26) with K given by (2.27) is inversely proportional to σ_v for small σ_v. Hence, this control selection method achieves the optimal worst-case information loss growth rate.

We next determine the optimal λ in (2.27) for the case where $v[n]$ is Gaussian with variance σ_v^2. We use $\hat{\mathcal{B}}_{\mathcal{N}}(x;\;\chi,\;K)$ to denote the Cramér-Rao bound (2.24) for $A = x\Delta$ in order to make its dependence on χ and on the period K in (2.26) explicit. The optimality of (2.27) suggests that K_{opt} from (2.25) is a non-decreasing function of χ for χ large. Indeed, there is a sequence χ_k where $k \ge 3$, such that, $K_{\text{opt}}(\chi) = k$, if $\chi_k < \chi < \chi_{k+1}$. If $\chi = \chi_k$,

both $K = k$ and $K = k+1$ minimize (2.25), *i.e.*,

$$\sup_{x\in(-1,1)} \hat{\mathcal{B}}_{\mathcal{N}}(x;\, \chi_k,\, k) = \sup_{x\in(-1,1)} \hat{\mathcal{B}}_{\mathcal{N}}(x;\, \chi_k,\, k+1)\ . \tag{A.14}$$

For large k the left hand side of (A.14) is maximized at $x = 1$ (*i.e.*, $A = \Delta$ in (2.24)), while the right hand side is maximized at $x = 1 - d(\chi_k; k)/2$ with $d(\cdot;\cdot)$ given by (2.28). Assuming that $d_{\text{opt}}(\chi)$ in (2.29) converges for large χ to a limit, *i.e.*, that $d_\infty = \lim_{\chi\to\infty} d_{\text{opt}}(\chi)$ exists, (A.14) reduces to

$$\sum_{n=-1}^{\infty} [\mathcal{B}_{\mathcal{N}}((n+1/2)\, d_\infty;\ 1)]^{-1} = \sum_{n=0}^{\infty} [\mathcal{B}_{\mathcal{N}}(n\, d_\infty;\ 1)]^{-1}\ , \tag{A.15}$$

where $\mathcal{B}_{\mathcal{N}}(A;\ \sigma)$ denotes $\bar{\mathcal{B}}(A;\ \sigma)$ for $v[n]$ Gaussian, and is given by (2.16) for $\sigma_\alpha = \sigma$. Both infinite series in (A.15) are convergent; in fact, only a few terms of each series are required to obtain an accurate estimate of d_∞ such as the one given in (2.31). Using d_∞ from (2.31) in conjunction with (A.14) and (2.24) yields (2.32). Similar results hold for nonGaussian sensor noise PDFs. Specifically, a relation of the form (A.15) holds for d_∞ defined in (2.30), where $\mathcal{B}_{\mathcal{N}}(\cdot\,;\ \cdot)$ is replaced by the associated $\bar{\mathcal{B}}(\cdot\,;\cdot)$. The resulting infinite series in (A.15) are both convergent since their terms decay faster than $1/n^\rho$ for $1 < \rho < 2$ (recall that $\mathcal{B}(A;\ y)$ grows faster than A^ρ for the control-free scenario). Clearly, the value of d_∞ depends on the particular noise PDF.

Extensions of the preceding control selection strategies can be developed, which achieve the optimal growth rate of the worst-case information loss for finite N. Let $\bar{\mathbf{w}}^N$ denote the control vector associated with the finite-N strategy, which is assumed known for estimation. Given a set of $w[n]$ and K selected according to infinite-N scheme, a finite-N method that achieves the same information loss for any A selects $\bar{\mathbf{w}}^N$ randomly from a set of K equally-likely vectors $\mathcal{W}(N,K) = \{\mathbf{w}_i^N :\ 1 \le i \le K\}$, where the nth element of the $N \times 1$ vector $\mathbf{w}_i^N$ is given by $w_i[n] = w[i\,N + n]$.

A.3 Information Loss for Signal Quantizers with $M \to \infty$

We consider a uniform quantizer with $M = 2(K+1)$ levels. Given K, we select the quantizer thresholds as $X_k = k/\sqrt{K}x$, where $k = -K, \cdots, K$, and $x > 0$. For convenience, we let $X_{-K-1} = -\infty$ and $X_{K+1} = \infty$. We next examine the Cramér-Rao bound (2.12) for

$w[n] = 0$, where $v[n]$ is admissible. We may rewrite (2.12) as

$$\mathcal{B}\left(A;\ \mathbf{y}^N\right) = \frac{1}{N}\left[\sum_{k=-K}^{K+1} \xi_k\right]^{-1} \tag{A.16a}$$

where

$$\xi_k = \frac{\left[p_v\left(X_k - A\right) - p_v\left(X_{k-1} - A\right)\right]^2}{C_v\left(X_{k-1} - A\right) - C_v\left(X_k - A\right)}\,. \tag{A.16b}$$

Note that as $K \to \infty$, both $\xi_{-K} \to 0$ and $\xi_{K+1} \to 0$. By letting $m_k = (X_k + X_{k-1})/2 - A$, for large K and for $k = -K+1, \cdots, K$ we have

$$\begin{aligned} p_v\left(X_k - A\right) - p_v\left(X_{k-1} - A\right) &\approx p_v'\left(m_k\right) x/\sqrt{K}\ , \\ C_v\left(X_{k-1} - A\right) - C_v\left(X_k - A\right) &\approx p_v\left(m_k\right) x/\sqrt{K}\ , \end{aligned}$$

which imply that

$$\xi_k \approx \frac{\left[p_v'\left(m_k\right)\right]^2}{p_v'\left(m_k\right)}\,\frac{x}{\sqrt{K}}\,. \tag{A.17}$$

Approximation (A.17) becomes an equality as $K \to \infty$. Letting $K \to \infty$ in (A.16) and using (A.17) yields

$$\begin{aligned} \lim_{K\to\infty} \mathcal{B}\left(A;\ \mathbf{y}^N\right) &= \frac{1}{N}\left[\lim_{K\to\infty}\left(\xi_{-K} + \xi_{K+1}\right) + \lim_{K\to\infty}\sum_{k=-K+1}^{K} \xi_k\right]^{-1} \\ &= \frac{1}{N}\left[\int_{t=-\infty}^{\infty} \frac{\left[p_v'\left(t - A\right)\right]^2}{p_v\left(t - A\right)}\ dt\right]^{-1} \\ &= \frac{1}{N}\left[\int_{t=-\infty}^{\infty} \left(\frac{\partial\ \ln\ p_v\left(t - A\right)}{\partial\, A}\right)^2 p_v\left(t - A\right)\ dt\right]^{-1} \\ &= \mathcal{B}\left(A;\ \mathbf{s}^N\right)\ . \end{aligned}$$

A.4 Asymptotic Efficiency of ML Estimator for the Case $M = 2$

In this appendix we show that $\hat{A}_{\mathrm{ML}} = \hat{A}_{\mathrm{ML}}\left(\mathbf{y}^N; \Delta\right)$ given by (2.40)–(2.42) achieves (2.14) for N large, if $\alpha[n]$ is admissible. Let $\tilde{k}$ denote the binomial random variable $\tilde{k}(\mathbf{y}^N) =$

$\mathcal{K}_1\left(\mathbf{y}^N\right)/N$. Then,

$$\hat{A}_{\mathrm{ML}} = \begin{cases} -\Delta & \text{if } \tilde{k} \leq C_\alpha(\Delta) \\ g(\tilde{k}) = -C_\alpha^{-1}\left(\tilde{k}\right) & \text{if } C_\alpha(\Delta) < \tilde{k} < C_\alpha(-\Delta) \\ \Delta & \text{if } \tilde{k} \geq C_\alpha(-\Delta) \end{cases} . \tag{A.18}$$

For large N the following approximation is valid in the cumulative sense

$$\tilde{k} \sim \mathcal{N}(p, \tilde{\sigma}_N), \tag{A.19}$$

where $p = C_\alpha(-A)$ and $\tilde{\sigma}_N = \sqrt{p(1-p)/N}$. Since $g(\cdot)$ is invertible ($C_\alpha(\cdot)$ is strictly monotone almost everywhere), the PDFs of $\hat{A}_{\mathrm{ML}}$ and $\tilde{k}$ are related as follows [26]

$$p_{\hat{A}_{\mathrm{ML}}}\left(\hat{A}\right) \approx \begin{cases} \delta(\hat{A}-\Delta)\, Q\left(\sqrt{N}\,\beta_+\right) & \text{if } \hat{A} = \Delta \\ p_{\tilde{k}}\left(C_\alpha\left(-\hat{A}\right)\right) p_\alpha\left(-\hat{A}\right) & \text{if } -\Delta < \hat{A} < \Delta \\ \delta(\hat{A}+\Delta)\, Q\left(\sqrt{N}\,\beta_-\right) & \text{if } \hat{A} = -\Delta \end{cases} , \tag{A.20}$$

where

$$\beta_+ = \frac{C_\alpha(-\Delta) - p}{\sqrt{p(1-p)}}, \qquad \beta_- = \frac{p - C_\alpha(\Delta)}{\sqrt{p(1-p)}} . \tag{A.21}$$

Note that the PDF of $\hat{A}_{\mathrm{ML}}$ in (A.20) consists of a sum of Kronecker delta functions.

We first consider $p_{\hat{A}_{\mathrm{ML}}}\left(\hat{A}\right)$ for $|\hat{A}| < \Delta$. If N is large enough, so that (A.19) is valid and also $\tilde{\sigma}_N \ll \sigma_\alpha$, the following approximations for $p_{\hat{A}_{\mathrm{ML}}}\left(\hat{A}\right)$ are valid in the regime $(-\Delta, \Delta)$, in the sense that for any A in $(-\Delta, \Delta)$ the values of the corresponding cumulative distribution functions are approximately equal (and where the approximation generally

improves as N increases)

$$p_{\hat{A}_{\mathrm{ML}}}(\hat{A}) \approx \frac{1}{\sqrt{2\pi}\tilde{\sigma}_N} \exp\left(-\frac{\left(C_\alpha\left(-\hat{A}\right)-C_\alpha(-A)\right)^2}{2\tilde{\sigma}_N^2}\right) p_\alpha\left(-\hat{A}\right) \quad \text{(A.22a)}$$

$$\approx \frac{1}{\sqrt{2\pi}\sqrt{\tilde{\sigma}_N^2\left(p_\alpha(-A)\right)^{-2}}} \exp\left(-\frac{\left(C_\alpha\left(-\hat{A}\right)-C_\alpha(-A)\right)^2}{2\tilde{\sigma}_N^2}\right) \quad \text{(A.22b)}$$

$$\approx \frac{1}{\sqrt{2\pi}\left(\gamma/\sqrt{N}\right)} \exp\left(-\left(p_\alpha(-A)\right)^2\frac{\left(\hat{A}-A\right)^2}{2\tilde{\sigma}_N^2}\right) \quad \text{(A.22c)}$$

$$\approx \frac{1}{\sqrt{2\pi}\left(\gamma/\sqrt{N}\right)} \exp\left(-\frac{\left(\hat{A}-A\right)^2}{2\left(\gamma/\sqrt{N}\right)^2}\right), \quad \text{(A.22d)}$$

where

$$\gamma^2 = C_\alpha(A)\; C_\alpha(-A)\; \left(p_\alpha(-A)\right)^{-2}\;.$$

Approximation (A.22a) results from using (A.19) in (A.20). To verify (A.22b) note that in the region that $\exp(-[C_\alpha(-\hat{A}) - C_\alpha(-A)]^2/[2\tilde{\sigma}_N^2])$ is essentially non-zero, we have $p_\alpha\left(-\hat{A}\right) \approx p_\alpha(-A)$. For $\tilde{\sigma}_N \ll \sigma_\alpha$, the following approximation is valid for the exponent in (A.22b)

$$\left[C_\alpha\left(-\hat{A}\right)-C_\alpha(-A)\right]^2 \approx \left(p_\alpha(-A)\right)^2\;\left(\hat{A}-A\right)^2\;,$$

which when substituted in (A.22b) results in (A.22c), and (A.22d). From (A.22d), for large N we have $\hat{A}_{\mathrm{ML}} \sim \mathcal{N}(A, \gamma^2/N)$ in the regime $(-\Delta, \Delta)$. Provided N is large enough, $\gamma^2/N \ll \Delta - |A|$, in which case the MSE term contributed from $\hat{A}_{\mathrm{ML}} \in (-\Delta, \Delta)$ approaches the Cramér-Rao bound γ^2/N. Next, consider the two other regimes, where $\hat{A}_{\mathrm{ML}} = \pm\Delta$. Let $\rho_- = \exp(-\beta_-^2/2)$, and $\rho_+ = \exp(-\beta_+^2/2)$, where β_- and β_+ are given by (A.21). For large enough N, $Q\left(\sqrt{N}\beta_+\right) \approx c_1\,\rho_+^N/\sqrt{N}$, and $Q\left(\sqrt{N}\beta_-\right) \approx c_2\,\rho_+^N/\sqrt{N}$. Since $0 < \rho_+, \rho_- < 1$, the corresponding MSE terms go to zero much faster than $1/N$ for large N, so their contribution to the MSE becomes negligible for large N as compared to γ^2/N. Hence, $\hat{A}_{\mathrm{ML}}$ achieves the Cramér-Rao bound (2.12) for large N.

A.5 EM Algorithm for Parameter Estimation in Gaussian noise via Signal Quantizers

In this appendix we present the derivation of an EM algorithm that can be used to obtain the ML estimate of an unknown parameter from a network of signal quantizers. The ith observation y_i is given by

$$y_i = F_i(x_i) \qquad i = 1, 2, \cdots, I \ ,$$

where

$$x_i = A + v_i + w_i \ , \tag{A.23}$$

A is the unknown parameter of interest, the sequence v_i is IID with $v_i \sim \mathcal{N}(0, \sigma_i^2)$, w_i is the selected (known) control input, and $F_i(\cdot)$ is the ith quantizer and is given by (2.2). We use $\underline{X}_i(\cdot)$ and $\overline{X}_i(\cdot)$ to denote the functions mapping each quantizer level Y_m of the ith quantizer $F_i(\cdot)$ to the associated lower and upper thresholds X_{m-1} and X_m, respectively.

We select as the complete set of data the set x_i in (A.23). For convenience, let

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_I \end{bmatrix}^T , \qquad \text{and} \qquad \mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_I \end{bmatrix}^T .$$

The EM algorithm selects $\hat{A}_{\text{EM}}^{(k+1)}$, the estimate of A at the $k+1$st step, based on $\hat{A}_{\text{EM}}^{(k)}$ and $\mathbf{y}$ according to

$$\hat{A}_{\text{EM}}^{(k+1)} = \underset{|\theta|<\Delta}{\arg\max}\ U\left(\theta;\ \hat{A}_{\text{EM}}^{(k)}\right) \tag{A.24}$$

where

$$U\left(\theta;\ \hat{A}_{\text{EM}}^{(k)}\right) = E\left[\ln p(\mathbf{x};\theta)|\mathbf{y};\hat{A}_{\text{EM}}^{(k)}\right] \ . \tag{A.25}$$

The log-likelihood function $\ln p(\mathbf{x};\theta)$ satisfies

$$\begin{aligned}\ln p(\mathbf{x};\theta) &= C - \sum_{i=1}^{I} \frac{1}{2\,\sigma_i^2}(x_i - w_i - \theta)^2 \\ &= h(\mathbf{x}) + \theta \sum_{i=1}^{I} \frac{1}{\sigma_i^2}(x_i - w_i) - \theta^2 \sum_{i=1}^{I} \frac{1}{2\,\sigma_i^2}\,. \end{aligned} \tag{A.26}$$

If we substitute the expression (A.26) for $\ln p(\mathbf{x};\theta)$ in (A.25) we obtain

$$U\left(\theta;\ \hat{A}_{\mathrm{EM}}^{(k)}\right) = E\left[h(\mathbf{x})|\mathbf{y};\hat{A}_{\mathrm{EM}}^{(k)}\right] + \theta\, E[k] - \frac{\theta^2}{2}\mu\,, \tag{A.27}$$

where $\mu = \sum_{i=1}^{I} \sigma_i^{-2}$, and

$$E[k] \;=\; \sum_{i=1}^{I} \frac{1}{\sigma_i^2} E_i[k] \;=\; \sum_{i=1}^{I} \frac{1}{\sigma_i^2}\left(E\left[x_i|\mathbf{y};\hat{A}_{\mathrm{EM}}^{(k)}\right] - w_i\right)\,. \tag{A.28}$$

Substituting in (A.24) the expression for $U\left(\theta;\ \hat{A}_{\mathrm{EM}}^{(k)}\right)$ in (A.27) we obtain

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(E[k]/\mu\right)\,. \tag{A.29}$$

Let $\underline{x}_i = \underline{X}_i(y_i)$ and $\overline{x}_i = \overline{X}_i(y_i)$. Using

$$p\left(x_i\,|\,\mathbf{y}\,;\,\hat{A}_{\mathrm{EM}}^{(k)}\right) \;=\; p\left(x_i\,|\,y_i\,;\,\hat{A}_{\mathrm{EM}}^{(k)}\right) \;=\; p\left(y_i\,|\,x_i\,;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\, p\left(x_i\,;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\, \left[p\left(y_i\,;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\right]^{-1}$$

we obtain

$$\begin{aligned} E_i[k] &= \frac{\frac{1}{\sqrt{2\,\pi}\sigma_i}\displaystyle\int_{x=\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}^{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i} \left(x+\hat{A}_{\mathrm{EM}}^{(k)}+w_i\right)\,\exp\left(-\frac{x^2}{2\,\sigma_i^2}\right)\,dx}{Q\left(\frac{\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right) - Q\left(\frac{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)} - w_i \\ &= \hat{A}_{\mathrm{EM}}^{(k)} + \frac{\sigma_i}{\sqrt{2\,\pi}}\,\frac{\exp\left(-\frac{\left(\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\,\sigma_i^2}\right) - \exp\left(-\frac{\left(\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\,\sigma_i^2}\right)}{Q\left(\frac{\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right) - Q\left(\frac{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)}\,, \end{aligned} \tag{A.30}$$

which when substituted in (A.29) results in

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)} + \frac{1}{\sqrt{2\,\pi}\,\sum_{i=1}^{I}\,\sigma_i^{-2}} \sum_{i=1}^{I} \frac{\exp\left(-\frac{\left(\underline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i\right)^2}{2\sigma_i^2} \right) - \exp\left(-\frac{\left(\overline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i\right)^2}{2\sigma_i^2} \right)}{\sigma_i \left[Q\left(\frac{\underline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i}{\sigma_i} \right) - Q\left(\frac{\overline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i}{\sigma_i} \right) \right]} \right) . \tag{A.31}$$

Several special cases of (A.31) are of interest. In particular, if $F_i(x) = F(x) = \mathrm{sgn}\,(x)$, (A.31) reduces to

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)} + \frac{1}{\sqrt{2\,\pi}\,\sum_{i=1}^{I}\,\sigma_i^{-2}} \left[\sum_{i=1}^{I} \frac{y_i}{\sigma_i} \frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{EM}}^{(k)} + w_i\right)^2}{2\sigma_i^2} \right)}{Q\left(-\frac{\hat{A}_{\mathrm{EM}}^{(k)} + w_i}{\sigma_i}\, y_i \right)} \right] \right) . \tag{A.32}$$

Next, consider the special case where N observations are collected from a single M-level quantizer (*i.e.*, $F_i(x) = F(x)$ and $I = N$). If, in addition, $w_i = w$ and $\sigma_i = \sigma$ for all i, (A.31) reduces to

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)} + \sum_{m=1}^{M} \frac{\sigma\,\mathcal{K}_{Y_m}(\mathbf{y})}{\sqrt{2\,\pi}\,N} \frac{\exp\left(-\frac{\left(X_{m-1} - \hat{A}_{\mathrm{EM}}^{(k)} - w\right)^2}{2\sigma^2} \right) - \exp\left(-\frac{\left(X_m - \hat{A}_{\mathrm{EM}}^{(k)} - w\right)^2}{2\sigma^2} \right)}{Q\left(\frac{X_{m-1} - \hat{A}_{\mathrm{EM}}^{(k)} - w}{\sigma} \right) - Q\left(\frac{X_m - \hat{A}_{\mathrm{EM}}^{(k)} - w}{\sigma} \right)} \right) ; \tag{A.33}$$

only the sufficient statistics $\mathcal{K}_{Y_1}(\mathbf{y})\,,\ \mathcal{K}_{Y_2}(\mathbf{y})\,,\ \cdots\,,\ \mathcal{K}_{Y_{M-1}}(\mathbf{y})$, are used in (A.33) to obtain $\hat{A}_{\mathrm{ML}}$.

A number of Generalized EM (GEM) algorithms can be derived which have interesting connections to the fast algorithms developed for parameter estimation in the presence of feedback. A GEM algorithm results in a sequence of estimates $\hat{A}_{\mathrm{GEM}}^{(k)}$ which have the property that at every step they increase $U\left(\theta;\ \hat{A}_{\mathrm{GEM}}^{(k)}\right)$ in (A.25) instead of maximizing it, *i.e.*,

$$U\left(\hat{A}_{\mathrm{GEM}}^{(k+1)};\ \hat{A}_{\mathrm{GEM}}^{(k)}\right) - U\left(\hat{A}_{\mathrm{GEM}}^{(k)};\ \hat{A}_{\mathrm{GEM}}^{(k)}\right) > 0\,. \tag{A.34}$$

Given the kth iteration estimate of this algorithm $\hat{A}_{\text{GEM}}^{(k)}$, the associated $U\left(\theta;\ \hat{A}_{\text{GEM}}^{(k)}\right)$ is given by (A.25), where $E[k]$ is given by (A.28)–(A.30) with $\hat{A}_{\text{EM}}^{(k)}$ replaced by $\hat{A}_{\text{GEM}}^{(k)}$, which we may rewrite for convenience as

$$E[k]/\mu = \hat{A}_{\text{GEM}}^{(k)} + \psi[k].$$

Consider the following class of iterative algorithms parameterized by λ

$$\hat{A}_{\text{GEM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(\hat{A}_{\text{GEM}}^{(k)} + \lambda\,\psi[k]\right)\ . \tag{A.35}$$

The algorithm corresponding to $\lambda = 1$ is the EM algorithm (A.31). Substituting $\hat{A}_{\text{GEM}}^{(k+1)}$ from (A.35) in (A.34) reveals (A.35) satisfies (A.34) if

$$\lambda(\lambda - 2) > 0\ .$$

Thus for $0 \leq \lambda < 2$ (A.35) yields a GEM algorithm. The algorithm corresponding to $\lambda = \pi/2$ is of particular importance, especially in the case $M = 2$ where feedback is present. In fact, it is the optimal λ when $\hat{A}_{\text{ML}} \approx 0$ in the case $M = 2$ and $w_i = 0$; the case $\hat{A}_{\text{ML}} = 0$ arises when $\mathcal{K}_1(\mathbf{y}) = I/2$. In this case, the convergence rate of the algorithm is given by

$$\lim_{\hat{A}_{\text{GEM}}^{(k)} \to \hat{A}_{\text{ML}}} \frac{\hat{A}_{\text{GEM}}^{(k+1)} - \hat{A}_{\text{ML}}}{\hat{A}_{\text{GEM}}^{(k)} - \hat{A}_{\text{ML}}} = 1 - 2\,\lambda/\pi\ .$$

From this point of view, the algorithm corresponding to $\lambda = \pi/2$ provides the optimal convergence rate when $\hat{A}_{\text{ML}}$ is close to the quantizer threshold. Consequently, it should not be surprising that its first step corresponds to the algorithm (2.58) when $n > n_o$, which was obtained heuristically and which achieves the optimal information loss in the Gaussian scenario for $M = 2$ in the context of feedback. In general, the GEM algorithm with $\lambda = \pi/2$ results in the ML estimate in fewer iterations than the EM algorithm for any set of observations, control input sequences, and noise levels.

The corresponding EM algorithms for MAP estimation can be readily derived if A is random with a priori distribution $\mathcal{N}(\mathrm{m}_A, \sigma_A^2)$. Specifically, we need only replace (A.24)

with [14]

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \arg\max_{\theta} \left[U\left(\theta;\ \hat{A}_{\mathrm{EM}}^{(k)}\right) + \frac{1}{2\,\sigma_A^2}(\theta - \mathrm{m}_A)^2 \right], \tag{A.36}$$

which results in the following MAP counterpart of (A.29)

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \frac{E[k] + \dfrac{\mathrm{m}_A}{\sigma_A^2}}{\mu + \dfrac{1}{\sigma_A^2}}, \tag{A.37}$$

where $E[k]$ is given by (A.28) and (A.30). MAP counterparts of (A.32) and (A.33) can be also readily derived.

A.6 Asymptotically Efficient Estimation for Pseudo-noise Control Inputs

In this appendix, we show that the estimators (2.45)–(2.47) of the parameter A presented in Section 2.3.1 are asymptotically efficient with respect to $\mathbf{y}^N$, where $y[n]$ is given by (2.1) with $F(\cdot)$ given by (2.2), and where $\alpha[n]$ in (2.10) is an IID admissible noise process. In the absence of pseudo-noise, $\alpha[n]$ equals $v[n]$. Consider the following collection of binary sequences

$$y_i[n] = F_i(A + \alpha[n]) = \mathrm{sgn}\,(A + \alpha[n] - X_i) \qquad i = 1, 2, \cdots, M-1\ .$$

The observed output $y[n]$ is equivalent to the collection $y_1[n], \cdots, y_{M-1}[n]$, since $y[n] = \sum_i y_i[n]$ and $y_i[n] = \mathrm{sgn}\,(y[n] - Y_i)$. The ML estimate of A based on $\mathbf{y}_i^N = \begin{bmatrix} y_i[1] & y_i[2] & \cdots & y_i[N] \end{bmatrix}^T$ is given by $\hat{A}_i$ in (2.45). We have

$$\hat{A}_i = \mathcal{I}_{\Delta}\left(-C_{\alpha}^{-1}\left(T_i\right)\right), \tag{A.38}$$

where

$$T_i \triangleq \mathcal{K}_1\left(\mathbf{y}_i^N\right) = \frac{1}{2\,N}\sum_n y_i[n] + \frac{1}{2}\ .$$

The estimators we develop next are based on the vector $\hat{\mathbf{A}}$ defined in (2.46). Note that, although the collection of T_i's is a set of sufficient statistics for the problem, $\hat{\mathbf{A}}$ is not, in general, a sufficient statistic due to the limiting operation in (A.38). In order to obtain the distribution of $\hat{\mathbf{A}}$ for large N, we need the distribution of the vector

$$\mathbf{T} = \begin{bmatrix} T_1 & T_2 & \cdots & T_{M-1} \end{bmatrix}^T .$$

For convenience, let $p_i \triangleq C_\alpha (X_i - A)$ and $f_i \triangleq p_\alpha (X_i - A)$. First note that the distribution of the vector $\left[\mathcal{K}_{Y_1} (\mathbf{y}^N) \; \mathcal{K}_{Y_2} (\mathbf{y}^N) \cdots \mathcal{K}_{Y_M} (\mathbf{y}^N) \right]^T$ is multinomial, and approaches a Gaussian vector in the cumulative sense [12]. The T_i's are linear combinations of the $\mathcal{K}_{Y_i} (\mathbf{y}^N)$'s, since $T_i = \sum_{j=i+1}^{M} \mathcal{K}_{Y_j} (\mathbf{y}^N)$. Consequently, $\mathbf{T}$ also approaches a Gaussian vector in the cumulative sense, *i.e.*, $\mathbf{T} \sim \mathcal{N}(\bar{\mathbf{T}}, R_{\mathbf{T}})$, where

$$\bar{\mathbf{T}} = \begin{bmatrix} p_1 & p_2 & \cdots & p_{M-1} \end{bmatrix}^T ,$$

$R_{\mathbf{T}} = R/N$, and

$$R = \begin{bmatrix} p_1 (1-p_1) & p_2 (1-p_1) & \cdots & p_{M-1} (1-p_1) \\ p_2 (1-p_1) & p_2 (1-p_2) & \cdots & p_{M-1} (1-p_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_{M-1} (1-p_1) & p_{M-1} (1-p_2) & \cdots & p_{M-1} (1-p_{M-1}) \end{bmatrix} .$$

In a manner analogous to the case $M = 2$ described in App. A.5, by using the theorem for the PDF of transformation of variables [26], and in the limit as $N \to \infty$ (where we invoke the law of large numbers and ignore the boundary effects due to $|\hat{A}_i| < \Delta$), we have $\hat{\mathbf{A}} \sim \mathcal{N}(A\,\mathbf{1}, C/N)$ in the cumulative sense, where $C = F^{-1} R F^{-1}$, and $F = \mathrm{diag}(f_1, f_2, \cdots, f_{M-1})$. It can be readily verified that

$$C^{-1} = \begin{bmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{M-2} \\ 0 & \cdots & 0 & b_{M-2} & a_{M-1} \end{bmatrix} , \tag{A.39}$$

where

$$a_i = \frac{f_i^2}{p_{i-1} - p_i} + \frac{f_i^2}{p_i - p_{i+1}} \qquad \text{and} \qquad b_i = \frac{f_i\, f_{i+1}}{p_{i+1} - p_i} \; .$$

If C where available, then the optimal estimate (in terms of minimizing the MSE) would be

$$\bar{A} = \boldsymbol{\lambda}^T \hat{\mathbf{A}} = \left(1^T C^{-1} 1\right)^{-1} 1^T C^{-1} \hat{\mathbf{A}} \; ,$$

while the associated MSE would satisfy

$$\lim_{N\to\infty} N\, E\left[(\bar{A} - A)^2\right] = \left(1^T C^{-1} 1\right)^{-1} = \mathcal{B}\,(A;\; y)$$

i.e., this estimator would be asymptotically efficient. However, C is a function of the unknown parameter A. Instead, note that $C(\hat{A}_i)$ approaches $C(A)$ for large N for any i (since $\hat{A}_i$ is asymptotically efficient). Specifically, set $i = 1$ and consider

$$\hat{A} = \boldsymbol{\lambda}(\hat{A}_1)^T \hat{\mathbf{A}} \; ,$$

where $\boldsymbol{\lambda}(\theta) = \left(1^T C^{-1}(\theta)\, 1\right)^{-1} 1^T C^{-1}(\theta)$. Let $\bar{z} = \bar{A} - A$, $z = \hat{z} - A$, and $\hat{\mathbf{z}} = \hat{\mathbf{A}} - A1$. Also, let $\Delta\boldsymbol{\lambda} = \boldsymbol{\lambda}(\hat{A}_1) - \boldsymbol{\lambda}(A)$, and denote by $\Delta\lambda_i$ the ith element of $\Delta\boldsymbol{\lambda}$. Then,

$$\begin{aligned}
\lim_{N\to\infty} N\, E\left[(\hat{A} - A)^2; A\right] &= \lim_{N\to\infty} N\, E\left[z^2; A\right] \\
&= \lim_{N\to\infty} N\, E\left[(\bar{z} + \Delta\boldsymbol{\lambda}^T \hat{\mathbf{z}})^2; A\right] \\
&= \mathcal{B}\,(A;\; y) + \lim_{N\to\infty} N \sum_{i,j} (\beta_{i,j} + \zeta_{i,j}) \, , \qquad \text{(A.40)}
\end{aligned}$$

where $\beta_{i,j} = E\left[\Delta\lambda_i\, \Delta\lambda_j\, \hat{z}_i\, \hat{z}_j\right]$, and $\zeta_{i,j} = E\left[\Delta\lambda_i\, \lambda_j\, \hat{z}_i\, \hat{z}_j\right]$. Note that in App. A.5 we have shown that $\lim_{N\to\infty} N\, E\left[(\hat{A}_1 - A)^2\right] = \mathcal{B}\,(A;\; y_1)$. Since $p_\alpha\,(\cdot)$ is admissible, for large N we have

$$\Delta\lambda_i \approx (\hat{A}_1 - A)\, \lambda_i'(A) \; .$$

Also, since $\hat{A}_1 - A$ is Gaussian for large N (see App. A.5), so is $\Delta\lambda_i$. In addition, there exists $G > |\lambda_i'(A)|$ for all i, which implies that $E\left[\Delta\lambda_i^2\right] \le G/N$, and $E\left[\Delta\lambda_i^4\right] \le 3\,G^2/N^2$. There also exists U such that $E\left[\hat{z}_i^2\right] \le U/N$ for all i, for N large enough. Finally, let

$\lambda_{\max} = \max_i \lambda_i(A)$. Repeated applications of the Schwarz inequality yield

$$|\beta_{i,j}| \leq \left(E\left[\Delta\lambda_i^4\right] E\left[\Delta\lambda_j^4\right] E\left[\hat{z}_i^4\right] E\left[\hat{z}_j^4\right]\right)^{1/4} \leq \frac{3GU}{N^2} ,$$

and

$$|\zeta_{i,j}| \leq \lambda_{\max}\left(E\left[\Delta\lambda_i^4\right] E\left[\Delta\lambda_j^4\right] E\left[\hat{z}_j^4\right]\right)^{1/4} \leq \frac{\lambda_{\max} 3^{3/4} G U^{1/2}}{N\sqrt{N}} ,$$

which, when substituted in (A.40), result in

$$\begin{aligned} \lim_{N\to\infty} N E\left[(\hat{A}-A)^2; A\right] &\leq \mathcal{B}(A;\, y) + \sum_{i,j} \lim_{N\to\infty} N\left(|\beta_{i,j}| + |\zeta_{i,j}|\right) \\ &\leq \mathcal{B}(A;\, y) + \sum_{i,j} \lim_{N\to\infty} \left(\frac{3GU}{N} + \frac{\lambda_{\max} 3^{3/4} G U^{1/2}}{\sqrt{N}}\right) \\ &\leq \mathcal{B}(A;\, y) \ . \end{aligned} \tag{A.41}$$

Since $\hat{A}$ is asymptotically unbiased (as a sum of asymptotically unbiased estimates), for N large we have $E\left[(\hat{A}-A)^2; A\right] \geq \mathcal{B}(A;\, y)/N$, which in conjunction with (A.41) yields the desired result (2.48).

Appendix B

B.1 EM Algorithm for Estimation of Gaussian Noise Parameters via Signal Quantizers

In this appendix we present the derivation of an EM algorithm that can be used to obtain the ML estimator of A and σ from a network of binary quantizers. The ith observation y_i is given by

$$y_i = F_i(x_i) = F_i(A + \sigma\, v_i + w_i) \qquad i = 1, \cdots, I \tag{B.1}$$

where A and σ are the unknown parameters of interest, satisfying $|A| < \Delta$ and $\underline{\sigma} < \sigma < \overline{\sigma}$, v_i is an IID sequence with $v_i \sim \mathcal{N}(0,1)$, w_i is a deterministic (known) sequence, and $F_i(\cdot)$ is the ith quantizer and is given by (2.2). We use $\underline{X}_i(\cdot)$ and $\overline{X}_i(\cdot)$ to denote the functions mapping each quantizer level Y_m of the ith quantizer $F_i(\cdot)$ to the associated lower and upper thresholds X_{m-1} and X_m, respectively.

We select as the complete data the set x_i in (B.1). For convenience we denote by $\mathbf{x}$, $\mathbf{y}$ the following vectors

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_I \end{bmatrix}^T$$

and

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_I \end{bmatrix}^T .$$

Let $\boldsymbol{\theta}$ denote the vector of parameters that we wish to estimate, *i.e.*,

$$\boldsymbol{\theta} \triangleq \left[A \quad \sigma \right]^T .$$

The EM algorithm then selects the $k+1$-st estimate $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k+1)}$ of $\boldsymbol{\theta}$ based on $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}$ according to

$$\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k+1)} = \underset{\boldsymbol{\theta}:\ |\theta_1| \leq \Delta}{\arg\max}\ U\left(\boldsymbol{\theta};\ \hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right) , \tag{B.2}$$

where

$$U\left(\boldsymbol{\theta};\ \hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right) = E\left[\ln p(\mathbf{x};\boldsymbol{\theta})|\mathbf{y};\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right] . \tag{B.3}$$

We have,

$$\ln p(\mathbf{x};\boldsymbol{\theta}) = h(\mathbf{x}) - I\,\ln(\sigma) + \frac{A}{\sigma^2}\sum_{i=1}^{I}(x_i - w_i) - \frac{I\,A^2}{2\,\sigma^2} - \frac{1}{2\,\sigma^2}\sum_{i=1}^{I}(x_i - w_i)^2 . \tag{B.4}$$

Substituting (B.4) in (B.3) and taking the derivative of $U\left(\boldsymbol{\theta};\ \hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right)$ with respect to A and σ, yields $\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k+1)}$, the next iteration estimate of $\boldsymbol{\theta}$, namely

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_{\Delta}\left(B[k]/I\right) \tag{B.5a}$$

$$\hat{\sigma}_{\mathrm{EM}}^{(k+1)} = \bar{\mathcal{I}}_{(\underline{\sigma},\,\overline{\sigma})}\left(\sqrt{\frac{G[k] + I\left[\hat{A}_{\mathrm{EM}}^{(k+1)}\right]^2 - 2\,\hat{A}_{\mathrm{EM}}^{(k+1)}\,B[k]}{I}}\right) , \tag{B.5b}$$

where

$$B[k] = \sum_{i=1}^{I} E\left[(x_i - w_i)|\mathbf{y};\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right] ,$$

$$G[k] = \sum_{i=1}^{I} G_i[k] ,$$

$$G_i[k] = E\left[(x_i - w_i)^2|\mathbf{y};\hat{\boldsymbol{\theta}}_{\mathrm{EM}}^{(k)}\right] ,$$

and where $\bar{I}_{(\underline{x},\,\overline{x})}(\cdot)$ is the following piecewise linear limiter function:

$$\bar{I}_{(\underline{x},\,\overline{x})}(x) = \begin{cases} x & \text{if } \underline{x} < x < \overline{x} \\ \underline{x} & \text{if } x \le \underline{x} \\ \overline{x} & \text{if } x \ge \overline{x} \end{cases} . \tag{B.6}$$

Letting

$$\underline{u}_i^{(k)} = \left(\underline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i\right)/\hat{\sigma}_{\mathrm{EM}}^{(k)}, \qquad \overline{u}_i^{(k)} = \left(\overline{x}_i - \hat{A}_{\mathrm{EM}}^{(k)} - w_i\right)/\hat{\sigma}_{\mathrm{EM}}^{(k)},$$

and

$$\underline{z}_i^{(k)} = \underline{x}_i + \hat{A}_{\mathrm{EM}}^{(k)} - w_i, \qquad \overline{z}_i^{(k)} = \overline{x}_i + \hat{A}_{\mathrm{EM}}^{(k)} - w_i,$$

the k-th samples of the sequences $B[k]$ and $G_i[k]$ are given by

$$B[k] = I\,\hat{A}_{\mathrm{EM}}^{(k)} + \frac{\hat{\sigma}_{\mathrm{EM}}^{(k)}}{\sqrt{2\,\pi}} \sum_{i=1}^{I} \frac{\exp\left(-\frac{\left[\underline{u}_i^{(k)}\right]^2}{2}\right) - \exp\left(-\frac{\left[\overline{u}_i^{(k)}\right]^2}{2}\right)}{Q\left(\underline{u}_i^{(k)}\right) - Q\left(\overline{u}_i^{(k)}\right)}, \tag{B.7}$$

and

$$G_i[k] = \left(\hat{A}_{\mathrm{EM}}^{(k)}\right)^2 + \left(\hat{\sigma}_{\mathrm{EM}}^{(k)}\right)^2 + \frac{\hat{\sigma}_{\mathrm{EM}}^{(k)}}{\sqrt{2\,\pi}} \frac{\exp\left(-\frac{\left[\underline{u}_i^{(k)}\right]^2}{2}\right) \underline{z}_i^{(k)} - \exp\left(-\frac{\left[\overline{u}_i^{(k)}\right]^2}{2}\right) \overline{z}_i^{(k)}}{Q\left(\underline{u}_i^{(k)}\right) - Q\left(\overline{u}_i^{(k)}\right)}, \tag{B.8}$$

which when substituted in (B.5), and in the limit $k \to \infty$ provide $\hat{A}_{\mathrm{ML}}(\mathbf{y})$ and $\hat{\sigma}_{\mathrm{ML}}(\mathbf{y})$. Note that when $\overline{z}_i^{(k)} = \infty$ in (B.8) (and thus also $\overline{u}_i^{(k)} = \infty$), we have

$$\exp\left(-\left[\overline{u}_i^{(k)}\right]^2/2\right) \overline{z}_i^{(k)} = \lim_{u\to\infty} \exp\left(-u^2/2\right)\left(u\hat{\sigma}_{\mathrm{EM}}^{(k)} + 2\hat{A}_{\mathrm{EM}}^{(k)}\right) = 0 .$$

Similarly, when $\underline{z}_i^{(k)} = -\infty$, we have $\exp\left(-\left[\underline{u}_i^{(k)}\right]^2/2\right) \underline{z}_i^{(k)} = 0$.

Appendix C

C.1 Proof of Theorem 1

In this section we prove the asymptotic optimality of the variable-rate encoding scheme described in Theorem 1. Using (4.11) and (4.9) we obtain

$$E\left[\left(\check{A}[N_{k+1}]-\hat{A}[N_{k+1}]\right)^2\right] < D\,M^{-2\,h[N_k]}\,, \tag{C.1}$$

where D is a constant satisfying $D \geq \Delta^2$. Using (C.1) and (4.10b) reveals that

$$\lim_{k\to\infty} N_k\,E\left[\left(\check{A}[N_{k+1}]-\hat{A}[N_{k+1}]\right)^2\right] = 0\,,$$

which in turn implies that

$$\lim_{k\to\infty} \frac{E\left[\left(\hat{A}[N_k]-A\right)^2\right]}{E\left[\left(\check{A}[N_k]-A\right)^2\right]} = 1\,. \tag{C.2}$$

Also, due to (4.11) and (4.10a) we have

$$\lim_{k\to\infty} \frac{E\left[\left(\hat{A}[N_{k+1}]-A\right)^2\right]}{E\left[\left(\hat{A}[N_k]-A\right)^2\right]} = 1\,,$$

which in conjunction with (C.2) implies (4.3), proving the asymptotic optimality of the algorithm.

C.2 Proof of Theorem 2

We will first prove the following lemma, which we will then use to prove Theorem 2.

Lemma 1 *Let*

$$\check{s}[n] = \text{sgn}\left(A - \check{A}[n]\right) .$$

Then the following statements are true for the dynamical system (4.16):

(a) *There exist arbitrarily large n such that*

$$\check{s}[n] = -\check{s}[n-1] . \tag{C.3}$$

(b) *If (C.3) is satisfied for some n', then*

$$|A - \check{A}[n]| < \frac{c}{n} . \tag{C.4}$$

(c) *The set of initial conditions for which (C.3) holds for all n satisfying $n > n'$ for some n' has zero measure, and satisfy*

$$\lim_{n\to\infty} |A - \check{A}[n]|^2 n^2 = \frac{c^2}{4} .$$

(d) *For almost all initial conditions there exist arbitrarily large n for which*

$$\check{s}[n] = \check{s}[n-1] = -\check{s}[n-2] . \tag{C.5}$$

Proof:

(a) To show (C.3) is satisfied for arbitrarily large n, we assume the opposite is true and arrive at a contradiction. Assuming there is an n' such that for all $n > n'$ $\check{s}[n] = \check{s}[n-1]$, and repeated use of (4.16) yield

$$\check{A}[n_b] > \sum_{k=n'}^{n} \frac{c}{k}$$

which must be true for all n. This, however, is a contradiction since $\sum_{k=n'}^{\infty} 1/k$ is not bounded for any n'.

(b) We can show this by induction. Due to (a) there exists an $n = n' + 1$ for which (C.3) is true. Since $\check{A}[n'+1]$ and $\check{A}[n']$ have opposite signs and satisfy (4.16), (C.4) is satisfied for n'. Assuming (C.4) holds for some $n > n'$ we show that it also holds for $n+1$. If $\check{s}[n] = -\check{s}[n-1]$ then (C.4) is satisfied for $n+1$. If on the other hand $\check{s}[n] = \check{s}[n-1]$, then since (C.4) holds for n

$$|A - \check{A}[n+1]| = |A - \check{A}[n]| - \frac{c}{n+1} < \frac{c}{n} - \frac{c}{n+1} < \frac{c}{n+1} .$$

(c) Let us assume that (C.3) is satisfied for all $n > n'$, where n' is even. Consider the sequence

$$x[n] = |A - \check{A}[n]| .$$

Then, for all $n > n'$, we have $x[n] > 0$ and also

$$x[n] = \frac{c}{n} - x[n-1] . \tag{C.6}$$

Repeated use of (C.6) and the fact that $x[n] > 0$ yields a relationship that must be satisfied for all even $n > n'$

$$\sum_{k=n'/2}^{n/2} \frac{1}{(2\,k)\,(2\,k+1)} < x[n'] < \frac{c}{n'} - \sum_{k=n'/2}^{n/2} \frac{1}{(2\,k+1)\,(2\,k+2)}$$

Since the limits (as $n \to \infty$) of both the upper and lower bounds on $x[n']$ above are equal, we must have

$$|A - \check{A}[n']| = \sum_{k=n'/2}^{\infty} \frac{1}{(2\,k)\,(2\,k+1)} . \tag{C.7}$$

Thus, (C.3) is satisfied for all $n > n'$ if and only if (C.7) holds. Finally, since (C.7) also holds for all even $n > n'$, we have

$$\lim_{k \to \infty} (2\,k)^2\, |A - \check{A}[2\,k]|^2 = \frac{c^2}{4}$$

The proof for odd n is similar.

(d) This is a trivial consequence of part (c).

■

The proof of Theorem 2 is a direct consequence of the above lemma. Specifically, (4.17) is trivially implied by part (b) of Lemma 1. In addition (C.4) implies that

$$\limsup_{n\to\infty} n^2 |A - \check{A}[n]|^2 \leq c^2 .$$

To show that c^2 is indeed the upper limit we employ part (d) of Lemma 1. Use of condition (C.5) reveals that there exist arbitrarily large n for which

$$n |A - \check{A}[n]| > c \left(1 - \frac{n}{(n-1)(n-2)}\right) ,$$

which completes the proof.

C.3 Asymptotic Optimality of the Digital Encoding and Estimation Algorithms of Sections 4.3.1–4.3.3

In this appendix we show the asymptotic optimality of the sequential encoder/estimator structures of the form (4.19) presented in Sections 4.3.1–4.3.3.

We first show the asymptotic efficiency of the fixed-rate encoder/estimator scheme of Section 4.3.1 consisting of (4.12) and (4.19). In the process we also derive the relationship (4.25) between β and λ. We will assume that (4.21) holds for large n and $n+1$, and find the value of β as $n \to \infty$. Since the resulting value is neither 0 nor ∞ the residual error indeed decays as $1/n^2$. By exploiting

$$\hat{A}[n+1] = \frac{n}{n+1} \hat{A}[n] + \frac{1}{n+1} s[n+1] ,$$

and (4.19b) we have

$$\begin{aligned} E\left[\left(\hat{A}[n+1] - \check{A}[n]\right)^2\right] = \mathrm{E}\Bigg[\Bigg(&\frac{n}{n+1}\left(\hat{A}[n] - \check{A}[n-1]\right) + \\ &\frac{1}{n+1}\left(s[n+1] - \check{A}[n-1]\right) - \\ &\frac{\lambda \sigma_v}{n} \operatorname{sgn}\left(\hat{A}[n] - \check{A}[n-1]\right)\Bigg)^2\Bigg] \end{aligned} \tag{C.8}$$

For convenience, let $\eta_1 = n\left(\hat{A}[n] - \check{A}[n-1]\right)/(n+1)$, $\eta_2 = \left(s[n+1] - \check{A}[n-1]\right)/(n+1)$,

and $\eta_3 = \lambda\,\sigma_v\,\mathrm{sgn}\,(\hat{A}[n] - \check{A}[n-1])/n$. Then, from (4.22)

$$\lim_{n\to\infty}\; n^2\,E\left[\eta_1^2\right] = \beta\,\sigma_v^2\,, \tag{C.9a}$$

and also

$$\lim_{n\to\infty}\; n^2\,E\left[\eta_3^2\right] = \lambda^2\sigma_v^2\,. \tag{C.9b}$$

The terms of the form $E\left[\eta_2\,\eta_i\right]$ for $i \neq 2$ decay faster than $1/n^2$, while

$$\lim_{n\to\infty}\; n^2\,E\left[\eta_2^2\right] = \sigma_v^2\,. \tag{C.9c}$$

The term in (C.8) corresponding to $E\left[\eta_1\,\eta_3\right]$ reduces to

$$E\left[\eta_1\,\eta_3\right] = \frac{\lambda\,\sigma_v}{n+1}\,E\left[\left|\hat{A}[n] - \check{A}[n-1]\right|\right]\,. \tag{C.10a}$$

Using (4.22) and the Schwarz inequality reveals that $E\left[\left|\hat{A}[n] - \check{A}[n-1]\right|\right]$ decays as $1/n$ or faster, *i.e.*,

$$\lim_{n\to\infty}\; n\,E\left[\left|\hat{A}[n] - \check{A}[n-1]\right|\right] = C \tag{C.10b}$$

where $0 \leq C < \infty$ is a function of λ, as is β. By multiplying both sides of (C.8) with n^2, taking the limit as $n \to \infty$ and by using (C.9) and (C.10b) we obtain

$$\beta\,\sigma_v^2 = \beta\,\sigma_v^2 + \sigma_v^2 + \lambda^2\,\sigma_v^2 - 2\lambda\,\sigma_v\,C\,. \tag{C.11}$$

If λ is chosen so that $0 < \lambda < \infty$, using (C.11) reveals that $0 < C < \infty$, which verifies that $E\left[\left|\hat{A}[n] - \check{A}[n-1]\right|\right]$ decays as $1/n$. This in turn implies that $E\left[(\hat{A}[n] - \check{A}[n-1])^2\right]$ cannot decay slower than $1/n^2$ (*i.e.*, $\beta > 0$). In fact, for large n, the PDF of $\hat{A}[n] - \check{A}[n-1]$ is well approximated by a Gaussian PDF in the cumulative sense. Consequently,

$$C(\beta) = \sqrt{2/\pi}\,\sqrt{\beta}\,\sigma_v\,, \tag{C.12}$$

which, when substituted to (C.11), results in (4.25).

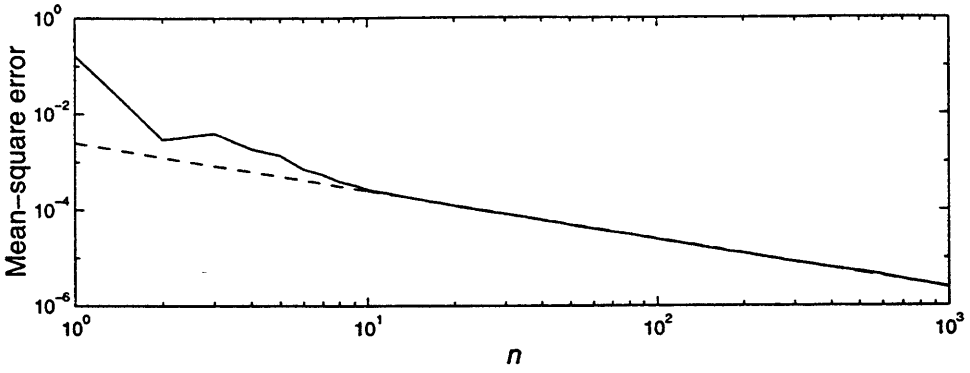
The accuracy of the Gaussian PDF approximation is examined in Fig. C-1, where we depict the results of Monte-Carlo simulations along with the associated predicted quantities from use of (4.25). Although this figure only shows the validity of the analysis for $\lambda = 1$, its accuracy is remarkable over a wide range of λ values that have been tested.

The analysis for the decoder/estimator system (4.12)–(4.19b) applies intact in the case that the noise is nonGaussian provided that $v[n]$ has finite variance, thus providing the asymptotic optimality of the encoder (4.12) and (4.19) in finite-variance sensor noises, described in Section 4.3.2.

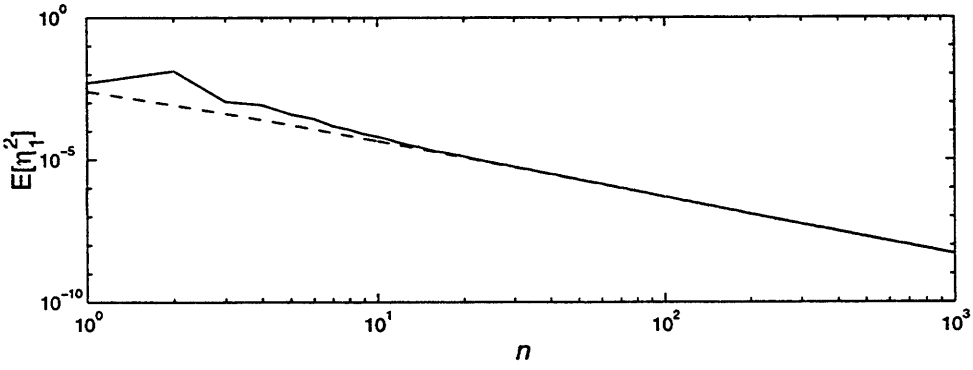
To show that the associated results hold for an encoder/estimator of the form (4.19) where $\hat{A}[n]$ is any estimator satisfying (4.27) and where the noise $v[n]$ is admissible, we again write

$$E\left[\left(\hat{A}[n+1]-\check{A}[n]\right)^2\right] = E\left[\left\{\left(\hat{A}[n]-\check{A}[n-1]\right)+\hat{\epsilon}[n+1]-\frac{\lambda\,\sigma_v}{n}\,\mathrm{sgn}\,\left(\hat{A}[n]-\check{A}[n-1]\right)\right\}^2\right] .$$

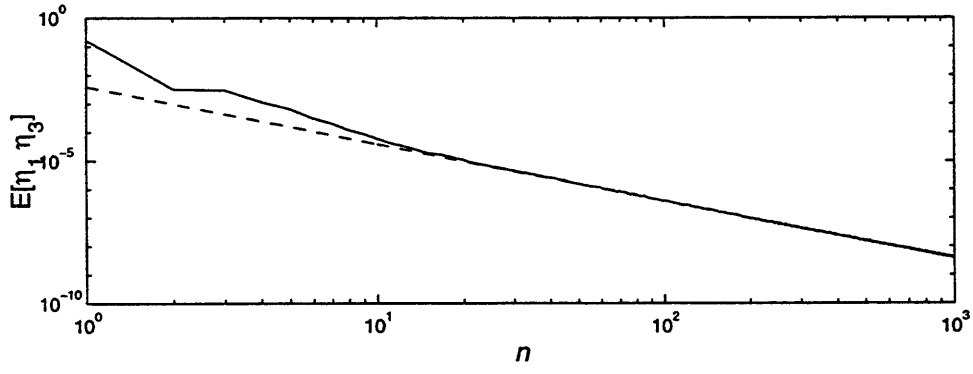
Similarly to the analysis of sample-mean based encoder, we make the association, $\eta_1 = \hat{A}[n]-\check{A}[n-1]$, $\eta_2 = \hat{\epsilon}[n+1]$, and $\eta_3 = \lambda\,\sigma_v\,\mathrm{sgn}\left(\hat{A}[n]-\check{A}[n-1]\right)/n$. The terms $E\left[\eta_1^2\right]$, $E\left[\eta_2^2\right]$, $E\left[\eta_3^2\right]$, and $E\left[\eta_1\,\eta_3\right]$ are given by (4.21), (4.27), (C.9b), and (C.10), respectively. Finally, we can easily show that the terms $E\left[\eta_2\,\eta_3\right]$ and $E\left[\eta_1\,\eta_2\right]$ also decay at least as fast as $1/n^2$ by using the fact that $E\left[\eta_i^2\right]$ decays as $1/n^2$ and the Schwartz inequality. All these together imply that the residual error decays as the reciprocal of the square of the observations. By ignoring the terms $E\left[\eta_2\,\eta_3\right]$ and $E\left[\eta_1\,\eta_2\right]$ we obtain an estimate for the scaling β of the residual error term in (4.29).



(a) MSE performance of $\check{A}[n]$ (solid) and Cramér-Rao bound $\mathcal{B}(A;\ \mathbf{s}^n)$ (dashed)



(b) $E\left[\left(\hat{A}[n]-\check{A}[n-1]\right)^2\right]$ vs. n



(c) $E\left[\left|\hat{A}[n]-\check{A}[n-1]\right|\right]$ vs. n

Figure C-1: Validity of the residual error analysis for $\lambda = 1$ for Gaussian $v[n]$. The solid lines on the lower two figures depict the results of Monte-Carlo simulations. The dashed curves correspond to the associated estimates obtained via the Gaussian approximation, leading to the value of β in (4.25). The dotted curve on the top-left figure denotes $\mathcal{B}(A;\ \mathbf{s}^n)$.

Bibliography

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, 1979.

[2] R. T. Antony. Database support to data fusion automation. *IEEE Trans. Signal Processing*, 85(1):39–53, January 1997.

[3] R. Benzi, A. Sutera, and A. Vulpiani. The mechanism of stochastic resonance. *J. Phys.*, A14:L453–L457, 1981.

[4] T. Berger, Z. Zhang, and H. Viswanathan. The CEO problem. *IEEE Trans. Inform. Theory*, 42(3):887–902, May 1996.

[5] R. S. Blum and S. A. Kassam. Optimum distributed detection of weak signals in dependent sensors. *IEEE Trans. Inform. Theory*, 38(3):1066–1079, May 1992.

[6] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part II— advanced topics. *IEEE Trans. Signal Processing*, 85(1):64–79, January 1997.

[7] D. A. Castanon and D. Teneketzis. Distributed estimation algorithms for nonlinear systems. *IEEE Trans. Automat. Contr.*, 30(5):418–425, May 1985.

[8] Z. Chair and P. K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerospace Elec. Sys.*, 22(1):98–101, January 1986.

[9] R. Chellappa, Q. Zheng, P. Burlina, C. Shekhar, and K. B. Eom. On the positioning of multisensor imagery for exploitation and target recognition. *IEEE Trans. Signal Processing*, 85(1):120–138, January 1997.

[10] P.-N. Chen and A. Papamarkou. Error bounds for parallel distributed detection under the Neyman-Pearson criterion. *IEEE Trans. Inform. Theory*, 41(2):528–533, March 1995.

[11] K. C. Chou, A. S. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans. Automat. Contr.*, 39(3):464–478, March 1994.

[12] H. Cramér. *Mathematical methods of statistics.* Princeton University Press, 1946.

[13] M. M. Daniel and A. S. Willsky. A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing. *IEEE Trans. Signal Processing*, 85(1):164–180, January 1997.

[14] A. P. Dempster, N.M Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Ann. Roy. Statist. Soc.*, 39:1–38, December 1977.

[15] M. DeWeese and W. Bialek. Information flow in sensory neurons. *Nuovo Cimento Soc. Ital. Fys.*, 17D(7-8):733–741, July–August 1995.

[16] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365:337–340, September 1993.

[17] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. Inform. Theory*, 37(2):269–275, March 1991.

[18] Z. Gingl, L. B. Kiss, and F. Moss. Non-dynamical stochastic resonance: theory and experiments with white and arbitrarily colored noise. *Europhys. Lett.*, 29(3):191–196, January 1995.

[19] E. B. Hall, A. E. Wessel, and G. L. Wise. Some aspects of fusion in estimation theory. *IEEE Trans. Inform. Theory*, 37(2):420–422, March 1991.

[20] S. Hein and A. Zakhor. Reconstruction of oversampled band-limited signals from $\sigma\delta$ encoded binary sequences. *IEEE Trans. Signal Processing*, 42(4):799–811, April 1994.

[21] S. Hein and A. Zakhor. Theoretical and numerical aspects of an SVD-based method for band-limited finite-extent sequences. *IEEE Trans. Signal Processing*, 42(5):1227–1230, May 1994.

[22] M. Kam, X. Zhu, and P. Kalata. Sensor fusion for mobile robot navigation. *IEEE Trans. Signal Processing*, 85(1):108–119, January 1997.

[23] J. Levin and J. Miller. Broadband neural encoding in the cricket sensory system enhanced by stochastic resonance. *Nature*, 380(6570):165–168, March 1996.

[24] Z.-Q. Luo and J. N. Tsitsiklis. Data fusion with minimal communication. *IEEE Trans. Inform. Theory*, 40(5):1551–1563, September 1994.

[25] D. Neuhoff, T. Pappas, and N. Seshadri. One-dimensional least-squares model-based halftoning. *J. Opt. Soc. Amer. A, Opt. and Image Sci.*, 14(8):1707–1717, August 1997.

[26] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, 3rd edition, 1991.

[27] T. Pappas. Digital halftoning: a model-based perspective. *Intern. J. Imag. Sys. and Technol.*, 7(2):110–115, February 1996.

[28] B. S. Rao, H. F. Durrant-Whyte, and J. A. Sheen. A fully decentralized multi-sensor system for tracking and surveillance. *Int. J. Robot. Res.*, 12(1):20–44, February 1993.

[29] N. T. Thao and M. Vetterli. Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates. *IEEE Trans. Signal Processing*, 42(3):519–531, March 1994.

[30] J. N. Tsitsiklis. Decentralized detection by a large number of sensors. *Math. Contr. Sig. Syst.*, 1(2):167–182, 1988.

[31] H. L. Van Trees. *Detection, Estimation and Modulation Theory, Part I.* John Wiley and Sons, New York, NY, 1968.

[32] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Inform. Theory*, 39(2):433–442, March 1993.

[33] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors: Part I—fundamentals. *IEEE Trans. Signal Processing*, 85(1):54–63, January 1997.

[34] D. Warren and P. Willett. Optimal decentralized detection for conditionally independent sensors. In *Amer. Control Conf.*, pages 1326–1329, 1989.

[35] M. Yeddanapudi, Y. Bar-Shalom, and K. R. Pattipati. IMM estimation for multitarget-multisensor air traffic surveillance. *IEEE Trans. Signal Processing*, 85(1):80–94, January 97.

[36] R. Zamir and M. Feder. Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization. *IEEE Trans. Inform. Theory*, 41(1):141–154, January 1995.